DEVELOPING A METHODOLGY TO DETERMINE

THE RELATIVE EFFECTIVNESS OF MALE AND HERMAPHRODITE PAPAYAS

IN CROSS-POLLINATION

A THESIS SUBMITTED TO THE GRADUATE DIVISION OF THE
UNIVERSITY OF HAWAI'I AT MĀNOA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

MOLECULAR BIOSCIENCES AND BIOENGINEERING

MAY 2017

By

Yosuke Kawai

Thesis Committee:

Richard Manshardt, chairperson
Ania Wieczorek
David Christopher

Keywords: qPCR, papaya, transgenic, DNA extraction

ACKNOWLEDGMENTS

I would first like to express my gratitude to the three committee members for the support and encouragement they gave me during this adventure. Over the course of the project, there were many instances in which I felt the project could not be completed due to technical difficulties and at each of these moments, they pushed me to be optimistic and believe that the hurdles can be overcome.

Dr. Richard Manshardt, thank you for the financial and mental support over the past five semesters. You were always logical, friendly, and willing to make many sacrifices for me. I greatly appreciate that you gave me a home during my stay here in the graduate program and at the same time, giving me an opportunity to work as your graduate assistant so that I can pay off my tuition. You have never scolded me once, even though there were many episodes where I deserved it!

Dr. Ania Wieczorek, thank you for inviting me to utilize your laboratory space to complete my project. Thank you for letting me borrow the quantitative PCR instrument and for letting me use the many reagents and supplies you have in the laboratory. Also, if it were not for your expertise in the technicalities in the qPCR method, there is no way that this project could have been completed.

Dr. David Christopher, thank you for being on my side of the entirety of my MBBE graduate program career. You have known me since the darkest of times in the beginning of the program and always gave me words of encouragement to move on. Your efforts in finding me a safe and productive laboratory to work in cannot be understated and I greatly appreciate your investment.

Dr. Jon-Paul Bingham, thank you for being an awesome graduate chair during the entirety of my stay in MBBE. I honestly think that you are probably the busiest graduate chair in the world, putting so much effort into making sure that EVERY SINGLE student in MBBE is working hard, but also while having a lot of fun. I put you in difficult situations many times in the past year and you never complained; you always took my side and put in the work so that I can work comfortably.

Dr. Kenneth Leonhardt, thank you for providing the male papaya leaf tissue whenever I needed it. I deeply regret that I was not able to thank you personally for your services. Who knew finding male papaya trees on O'ahu would be so hard?

Kauahi Perez and Dr. Angel Del Valle-Eschevarria, thank you so much with the technical support you provided over the past two years. To me, you guys were the "walking encyclopedia" and in hindsight, I think I was a spoiled graduate student to have someone like you guys support me throughout my stay.

I would like to give my special thanks to all my family members. Although we were separated by many, many miles, the words of support you gave me through Skype was the push I needed to get through graduate school.

I would also like to give my special thanks to all my friends that helped me get to this point. These great friends were always with me, in my brightest and darkest days. Being able to drink beers with you guys was a great way to recharge my motivation and keep my sanity.

Ms. Han Ha Youn, you deserve a special section of gratitude for all you have done for me. If I could, I would write everything you have done for me during our relationship, but then my thesis will end up being a 1,000 pages long. Thank you for your love, support, and your willingness to pick me up from school at 3 a.m.!

ABSTRACT

It has been almost two decades since the introduction of transgenic papaya saved the crop in Hawai'i from the devastating effects of the papaya ringspot virus. The virus-resistant papaya was rapidly adopted by farmers, and it now accounts for approximately 80% of papaya acreage in the State. Although transgenic fruits have passed USDA, EPA, and FDA testing to become an integral part of the papaya market, there are also concerns of uncontrolled transgene flow into organic production systems and crops bound for international markets. Pre-commercialization testing of new transgenic crops is becoming more stringent and researchers will need to have a better understanding of the crop's reproductive behavior, so they can correctly assess the biological risk and design appropriate bio-containment procedures. Provided with these data, there will be less room for false speculations regarding gene flow between plants. For papaya, it is currently known that gene flow by seed is more significant than by pollen, but this situation could change if male plants from feral dioecious (male and female) populations become transgenic by out-crossing with transgenic varieties. Male plants generally produce more flowers than hermaphrodites, which suggests that the role of pollen in transgene flow would increase. Currently, no data exist regarding the relative effectiveness of male and hermaphrodite papaya as pollen contributors to female and hermaphrodite papaya that receive the pollen. The purpose of this study is to design a qPCR methodology for detecting fingerprints of male and hermaphrodite pollen in seed batches and quantifying the percent contribution from the two pollen sources.

The genetic contribution by male papaya is tracked by amplifying a product from a male-specific marker (PMSM-1), whereas the hermaphrodite contribution is tracked using the

transgenic *coat protein* gene. The primer and probe sequences of the *coat protein* gene has been previously published, so the first objective of the thesis is to design primer and probes that amplify a product from PMSM-1. Primers and probes that bind specifically to the *male* marker was successfully developed. Next, a CTAB-based DNA extraction procedure that can yield high amounts of high quality DNA for qPCR usage was developed. Through the combination of the traditional CTAB method and the DNeasy Plant Mini Kit, the method could yield approximately 70 ng/µL of amplifiable DNA with a purity of 1.64 to 1.97 (260/280 absorbance ratio).

A reliable and sensitive qPCR assay for detecting and quantifying the *coat protein* gene and the *male* marker in bulked seed and leaf tissue samples was developed in this thesis. There was a strong linear relationship in the standard curve between the measured fluorescence and the total DNA concentration when the template concentration was 100 to 1 ng per reaction. The data were normalized using the $2^{-\Delta\Delta Ct}$ method, adjusting for variation in the amount of template DNA, provided that all primer efficiencies were operating at similar efficiencies. The sensitivity and reliability of the assay was validated by extracting DNA from known ratios of male DNA to hermaphrodite DNA, followed by qPCR detection and quantification.

The methodology outlined in this thesis is a reliable and sensitive assay allowing for the detection and accurate quantification of the pollen contribution by the male and hermaphrodite papayas. The results from this thesis can benefit future researchers by outlining a quick qPCR protocol in screening papaya seeds for different genetic markers. This will further add to the knowledge of the reproductive biology of papaya.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABI	Applied Biosystems, Inc.
ASGPB	Advance Studies in Genomics, Proteomics, and Bioinformatics
BAC	bacterial artificial chromosome
bp	base pairs
BLAST	basic local alignment search tool
cp	*coat protein*
cDNA	complementary DNA
CTAB	cetyl-trimethyl-ammonium-bromide
C_t	cycle threshold
°C	degrees Celsius
DNA	deoxyribonucleic acid
dNTP	deoxy-nucleotide triphosphates
$	dollars
dsDNA	double stranded DNA
EDTA	ethylene-diamine-tetra-acetic acid
EPA	Environmental Protection Agency
et al.	“and others”
FDA	Food and Drug Administration
FRET	fluorescence resonance energy transfer
g	gravitational force
GATA	guanine-adenine-thymine-adenine

GC	guanine-cytosine
GE	genetically engineered
GMO	genetically modified organism
HCl	hydrochloric acid
HSY	hermaphrodite specific Y
IDT	Integrated DNA Technologies, Inc.
in vitro	performed or taking place in a test tube
kcal	kilo calories
kg	kilo grams
M	molar
Mb	mega bases
MEGA	Molecular Evolutionary Genetics Analysis
$MgCl_2$	magnesium chloride
µL	micro liter
µM	micro molar
mL	milli liter
mM	milli molar
mol	moles
mRNA	messenger RNA
MSY	male specific Y
NaCl	sodium chloride
NaOH	sodium hydroxide
ng	nanograms

OH	hydroxide
%	percentage
pg	pico grams
PCR	polymerase chain reaction
pH	potential of hydrogen
PMSM	papaya male specific marker
PRSV	papaya ringspot virus
PVP	poly-vinyl-pyrrolidone
PVPP	poly-vinyl-poly-pyrrolidone
qPCR	quantitative polymerase chain reaction
R	correlation coefficient
R^2	coefficient of determination
RAPD	random amplified polymorphic DNA
RNA	ribonucleic acid
SCAR	sequence characterized amplified region
SNPs	single nucleotide polymorphisms
ssDNA	single stranded DNA
TBE	Tris-HCl, boric acid, EDTA
TE	Tris-HCl, EDTA
T_m	melting temperature
Tonnes	equivalent to 1,000 kilograms
US	United States
UV	ultraviolet

V	volts
VAD	vitamin A deficiency
WHO	World Health Organization
Y^h	hermaphrodite Y chromosome
Y^m	male Y chromosome

CHAPTER 1. INTRODUCTION

1.1 Papaya as a commodity

Carica papaya L., commonly known as papaya, is a perennial diploid (2*n*=18) tree in the Caricaceae plant family. The origin of wild papaya shows that they are native to Central America from southern Mexico to northern Costa Rica (Antunes Carvalho & Renner, 2012; Candolle, 1885). Taxonomic revisions have made papaya the lone member of the *Carica* species, and the idea of an interspecific hybrid origin for papaya is not supported by any data. There is successful commercialization of papaya in many of the tropical parts of the world and to a lesser extent, in subtropical regions, such as Australia and South Africa.

Papaya is one of the most traded tropical fruit in the world, with a global production of approximately 12.7 million tonnes in 2014 (FAOSTAT, 2014) with India producing nearly half of the total (Table 1). This same year, the United States produced 13,000 tonnes of papaya. In the past decade, the Hawaiian papaya market has been worth approximately $10 million, reaching up to $14 million in 2008 and 2009 (NASS, 2012).

Table 1. Top 5 producers of papaya in 2014 (FAOSTAT, 2014).

Country/State	Production (million tons)
India	**5.5**
Brazil	**1.6**
Indonesia	**0.9**
Nigeria	**0.8**
Mexico	**0.8**
World	**12.4**

Papaya fruit is the most nutritious of the 38 commonly consumed fruits based on percentage of US recommended daily allowance for compounds such as Vitamin A, Vitamin C, folate, potassium, iron, calcium, and fiber (Chandrika, Jansz, Wickramasinghe, & Warnasuriya, 2003). Consumption of papaya is often recommended to fight micronutrient deficiency, especially in developing countries. One example is the vitamin A deficiency (VAD), which is the leading cause of preventable blindness in children and can be a cause of death from severe infections ("WHO | Micronutrient deficiencies," n.d.). It is a common problem in Africa and South-East Asia, causing blindness in 250,000 to 500,000 children. Food fortification is one of the most common solution to VAD and papaya is one of the most recommended fruits and vegetables for this disorder.

1.2 History of Genetically engineered (GE) Papaya

Genetically engineered (GE) papaya was developed as a response to the devastating disease caused by papaya ringspot virus (PRSV). Genetically modified organisms, or GMOs, are organisms such as plants, animals, and bacteria in which the genetic material (DNA or RNA) is altered in a way that does not occur naturally. This alteration involves the mutation, insertion, or deletion of genetic material to introduce a "trait of interest" to an organism (Ahmed, 2002).

The PRSV is notorious for being the most widespread and damaging virus that infects papaya (Gonsalves, 1998). The name of the disease is derived from one of the symptoms shown on an infected papaya fruit (Figure 1). Other symptoms include: oily streaks on the petiole and upper part of the trunk, distortion of young leaves that resembles mite damage, and stunted tree growth (Gonsalves, 1998). The effects of the infection are most severe when the infection takes place at the seedling stage, after which the tree does not produce a mature fruit. The PRSV affects papaya crops worldwide.

The PRSV has affected Hawaiian crops since 1945, when the virus was first discovered (Jensen, 1949). The virus was first discovered on the island of O'ahu, and followed the industry as it moved to the Puna district on the Big Island. Although PRSV was eradicated by rapid roguing of trees in 1965, Dr. Dennis Gonsalves from Cornell University and researchers at the University of Hawai'i were convinced that the virus would inevitably invade Puna again. Collaboration with Taiwanese universities showed that cross protection, a method in which plant crops are systemically infected with a mild strain of a virus was able to delay the onset of the disease, however, this method did not offer complete protection (Yeh, 1988) against the disease. In the mid-1980s, a method called parasite-derived resistance was adopted, which was a way of artificially creating a viral gene and inserting it into the plant genome to protect the plant against

detrimental effects of similar pathogens. This methodology was first successfully utilized in tobacco to protect against infection from the tobacco mosaic virus (Fraley et al., 1983). Following a similar procedure, the Hawaii variant of the papaya ringspot virus (called PRSV HA 5-1) was cloned and sequenced. Specifically, the target gene of PRSV was the *coat protein* gene in PRSV HA 5-1. The papaya cultivars Sunrise, Sunset, and Kapoho were tested for the transformation process. The transformation process consisted of bombarding papaya embryogenic tissue with tungsten particles coated with the DNA of the *coat protein* gene using the gene gun.

The resulting transgenic plants were resistant to Hawaiian isolates of PRSV and mildly resistant to non-Hawaiian isolates. The transgenic Sunset cultivar, called line 55-1 during trials and later renamed as SunUp, is a homozygous transgenic line. At this time, the yellow-fleshed Kapoho was the dominant cultivar growing in Puna and this non-transgenic cultivar was crossed with SunUp, and the hybrid made from this cross was named UH Rainbow (Manshardt, 1999). These transgenic cultivars were deregulated by the FDA and EPA in 1997 and commercialized in Hawai'i in mid-1998.

Figure 1. Characteristic ringspot pattern on PRSV-infected papaya.

1.3 Purpose of Thesis

The introduction of transgenic papaya saved the papaya industry in Hawai'i from the devastating disease caused by PRSV, but there are growing concerns of unwanted transgene flow into organic fields which could lead to a decrease in exports. Although this is not the specific problem this thesis is going to address, it is a fact that there are inadequate data regarding the reproductive behavior of papaya and the biological risks associated with its transgenic variants. This thesis project serves to provide definitive data on one aspect of transgene flow, which is gene flow by pollen. The purpose of this thesis is to determine the relative effectiveness of male and hermaphrodite papayas in cross-pollinating other papayas. The author and the committee members are optimistic that these data will aid future researchers in designing appropriate bio-

containment protocols to minimize escape of transgenes during the pre-commercialization testing phase.

1.4 Objectives

The overall goal of this study is to develop an assay using qPCR technology to distinguish and quantify pollen contribution from hermaphrodite and male papayas. This assay will ultimately be used to determine the relative pollen contribution of these two sexes by quantifying sex-specific signals. After completion of this thesis project, the assay will be used to test the seeds harvest from an experimental plot (Figure 2) to determine the relative pollen contribution of hermaphrodite and male papayas. In this plot, transgenic hermaphrodite and non-transgenic male papayas (rows 28 to 30) are the pollen contributors, whereas the rest of the plot, consisting of non-transgenic female and non-transgenic hermaphrodite papayas are the pollen receivers.

In this thesis, the *coat protein* gene, which is a component of the transgene in GE papaya, will be used as a marker for hermaphrodite pollen. The sequences for primers and probes that amplify the *coat protein* gene has been previously published and showed good results (Nageswara-Rao et al., 2013; Xu et al., 2008). Markers specific to the male papaya genome has also been identified recently (Liao, Yu, & Ming, 2017), so the first objective of this thesis project is to design primers and probes that target this marker.

The second objective is to test and optimize a DNA extraction methodology for both papaya seed and leaf tissue. Existing procedures (Guo et al., 2009; Nageswara-Rao et al., 2013; Xu et al., 2008; Zhu et al., 2012), particularly the CTAB methods, will be tested and optimized.

The third objective is to optimize a qPCR methodology so that signals from the *coat protein* gene and the male-specific marker can be detected and quantified. Standard curves that relate the fluorescent signals detected by the instrument to the concentration of the DNA will be established and interpreted.

The fourth objective is to validate the standard curves by testing DNA mixtures with known proportions of male and hermaphrodite templates. The resulting qPCR fluorescent signals will be normalized and compared to expected values generated by the standard curve.

Waimanalo T Field Planted July 23, 2012; males & GMO herm. planted Oct. 29, 2012

Plant #	Row 1	Row 2	Row 3	Row 4
1	F	F	F	F
2	F	F	F	F
3	F	F	F	F
4	h	F	F	F
5	h	F	h	F
6	F	F	F	F
7	h	F	h	F
8	F	h	F	F
9	F	F	F	F
10	h	F	h	F
11	F	F	F	F
12	F	h	F	h
13	F	h	F	h
14	h	F	F	F
15	F	F	F	F
16	F	F	F	F
17	F	h	F	h
18	h	F	h	F
19	h	F	h	F
20	F	F	F	F
21	F	h	F	h
22	h	F	F	F
23	F	h	F	F
24	F	F	F	F
25	h	F	h	F
26	F	h	F	F
27	F	h	F	h
28	M	H	M	H
29	H	M	H	M
30	M	H	M	H

Fruits stripped Oct. 29, 2012

Female

Hermaphrodite

GMO Hermaphrodite

Male

Date	Flagging color
11/16/12	Black
11/28/12	wh/org
12/5/12	blue/white

Figure 2. Diagram of the experimental plot at the Waimanalo Research Station. The designated pollen sources in this experiment are the transgenic hermaphrodites and non-transgenic males in rows 28 to 30. The rest of the plants are non-transgenic.

CHAPTER 2. LITERATURE REVIEW

2.1 Differences between the papaya sexes

2.1.1 Introduction

The papaya sex system is polygamous with three known sex types: female, male, and hermaphrodite. The wild populations of papaya are generally dioecious (male and female), whereas gynodioecious systems arose recently as a product of domestication (Vanburen et al., 2015). Sex determination in papaya is controlled by a pair of recently evolved sex chromosomes with XX genotype as female, XY^m male and XY^h hermaphrodite (Ma et al., 2004; Zhiyong et al., 2004). In the wild, female plants cross with males plants, to produce a plant that segregates in a 1:1 ratio (Ming, Yu, & Moore, 2007) of the parent sexes. With the gynodioecious system seen in commercial fields, two types of crosses occur, either hermaphrodite trees self-pollinate or hermaphrodite plants pollinate the female. Self-pollinated hermaphrodite plants always segregate in a 2 hermaphrodite to 1 female ratio, in which homozygosity for the Y sex chromosome is a lethality factor. Crosses between female and hermaphrodite plants segregate in a 1:1 ratio of the parent sexes, similarly to the dioecious system (Table 2).

Table 2. Three different crossing scenarios of papaya sexes.

Parent Mating	**Progeny Sex Ratios**			
	Male (XY^m)	**Hermaphrodite (XY^h)**	**Female (XX)**	**Lethal (YY)**
Male (XY^m) x Female (XX)	**1**	**0**	**1**	**0**
Herm (XY^h) x Female (XX)	**0**	**1**	**1**	**0**
Hermaphrodite (XY^h) selfed	**0**	**2**	**1**	**1**

2.1.2 Physical traits of the papaya sexes

The papaya plant shows no sexual dimorphism prior to flowering and once they flower, differences in the physical traits begin to show. Hermaphrodite and female plants have short inflorescences with few flowers, in contrast to panicle of male plants which contain dozens of flowers and are very long and pendulous (Decraene & Smets, 1999). Papaya fruit also varies in shape and size. Hermaphrodite plants produce fruits that are elongated or pear-shaped, while those of female plants tend to be more rounded or oval (Paull, Nishijima, Reyes, & Cavaletto, 1997). The fruit mass can vary from 0.3 kg to as heavy as 10 kg (Paull et al., 1997), depending on the variety. Typically, hermaphrodite trees are more desirable producers because their ability to self-pollinate eliminates the need of growing non-fruiting pollen donors (Rieger, 2006).

2.1.3 Genetic characteristics of papaya

Papaya is the first genetically modified crop to have its entire genome sequenced. The papaya genome consists of 372 mega base pairs (Ming et al., 2008) and is considered one of the model plants to study because the plants are relatively easy to maintain. The papaya genome

consists of 9 pairs of diploid chromosomes, with chromosome pair 1 are the sex chromosomes. The papaya sex chromosomes are unique in that it has two variants of the Y chromosome, distinguished by a small sex-determining region that is suppressed for recombination (Chen et al., 2007; Ma et al., 2004).

With the evolution of molecular biology, mapping sex-linked markers in papaya became possible. In 1999, a microsatellite sequence specific to male and hermaphrodite papayas were identified (Parasnis, Ramakrishna, Chowdari, Gupta, & Ranjekar, 1999). A few years later, two male and hermaphrodite sequence characterized amplified region (SCAR) markers were developed (Deputy et al., 2002) and used to construct the physical map of a hermaphrodite-specific region on the Y^h chromosome (HSY) via chromosome walking (Na et al., 2012; Wang et al., 2012). Other sex-linked SCAR markers have also been developed, but they were specific for both hermaphrodites and males (Parasnis, Gupta, Tamhankar, & Ranjekar, 2000). Sequencing the papaya sex-specific regions of the two Y chromosomes revealed that HSY and the male-specific region of the Y^m chromosome (MSY) had 4.6 Mb extra DNA sequences compared to their X counterpart (Na et al., 2012; Wang et al., 2012). This explained why sex-linked RAPD and SCAR markers that distinguish males and hermaphrodites from females are easy to identify and develop. However, no markers that distinguish the male and the hermaphrodite have been identified due to the high level of sequence identify between MSY and HSY, which was 99.6% (Vanburen et al., 2015). It is theorized that the Y^h was derived from reverse mutation of the carpel suppressing gene on the male Y chromosome during the domestication process of papaya 4,000 years ago (Vanburen et al., 2015).

Later, Liao et al. (2017) used a large male-specific retrotransposon insertion of 8,396 bp to identify two papaya male-specific markers, PMSM-1 and PMSM-2. The primers designed for

PMSM-1 targeted a sequence between positions 7,270 and 7,854 bp to yield a 585 bp fragment. Likewise, the primers designed for PMSM-2 targeted a sequence between positions 3,566 to 4,113 bp to yield a 548 bp fragment. To the author's knowledge, this is the first manuscript that described the successful development of PCR primers that target a male-specific sequence.

2.2 Polymerase Chain Reaction (PCR)

The polymerase chain reaction, or PCR, is an important molecular biology technique used for diagnostic and monitoring research. Developed by Dr. Kary Mullis in 1983 (Mullis et al., 1986), this technique utilizes an enzymatic chemical reaction to amplify a few copies of DNA, *in vitro*, up several orders of magnitude to generate thousands of copies of the amplicon. This technique has been used for a variety of research, including disease diagnosis, forensics, assay development, DNA cloning, and phylogeny.

The basic components of the PCR are the DNA template, primers, deoxynucleoside triphosphates (dNTPs), polymerase, buffer, and bivalent cations. The template is usually double-stranded DNA (dsDNA) or synthesized single-stranded complementary DNA (ss-cDNA) that contains the target sequence of interest. For successful PCR, the user needs to be aware of the sequence of the target DNA, so that sequence-specific oligonucleotide primers that hybridized to the target can be designed. These primers are short, ssDNA fragments synthesized to have a complementary sequence to the target DNA. During PCR, these fragments bind to the single-stranded target, with the 3' end hydroxide (-OH) group left exposed for the addition of nucleosides. dNTPs include deoxyadenosine, deoxyguanosine, deoxycytidine, and deoxythymidine, which are the four bases that make up the DNA structure. These nucleosides are added to the free 3' end of the primers one by one by a polymerase, an enzyme that

assembles the nucleoside to create a strand of DNA that has a complementary sequence to the target. Most reactions utilize a heat-stable DNA polymerase, most commonly the *Taq* polymerase, named after the thermophilic bacterium *Thermus aquaticus*, where the enzyme was first isolated from (Brock & Freeze, 1969). The enzyme operates optimally at 70°C. The components described above are combined in specific proportions into the buffer solution, which keeps the pH in the solution nearly constant during the chemical reaction. Finally, bivalent cations, most commonly magnesium ions, are needed to remove the phosphate groups of the dNTPs, so that the polymerase can continue adding new nucleosides.

The PCR method utilizes thermal cycling, which are repeated cycles of heating and cooling of the reaction to specific temperatures (Figure 3). PCR typically consists of 35 to 45 cycles (repetitions) of three different thermal steps: denaturation, annealing, and extension. In an ideal PCR, the amount of template doubles each cycle. The reaction is first heated to 95°C to denature, or separate, the dsDNA template into two ssDNA by disrupting the hydrogen bonds between complementary bases of the double-stranded template. The temperature is then lowered to about 50 to 65°C to allow the primers to hybridize to the ssDNA templates. This temperature is optimized depending on the characteristics of the primer sequence. Ideally, this annealing temperature should be set low enough to allow for proper primer hybridization to the template, but high enough so the primers do not anneal back to one another. In the extension step, the reaction is heated to 70-75°C, allowing the DNA polymerase to add dNTPs to the primer sequence in a 5' to 3' direction. Recently, some DNA polymerases used in PCR require a "hot-start", activating only after the enzyme is exposed to a certain temperature. This additional step, often called the initiation step, takes place 3 to 5 minutes prior to the first denaturation step. The main advantage of this polymerase is that it is unreactive at ambient temperatures, which allows

the convenience of setting up the reaction at room temperature, while reducing non-specific amplification to a minimum.

The basic PCR described above is also known as end-point PCR because the analysis of the products is performed after the completion of the PCR cycles. The quality and quantity of the products are typically determined using an agarose gel. This type of analysis is susceptible to variation between reactions and is difficult to provide accurate quantitative information of the starting sample. Thus, this type of analysis is performed using a real-time PCR system.

2.3 Real-time PCR

2.3.1 General Introduction

Real-time, or quantitative, PCR is a technique that monitors the amplification of the target DNA during the PCR. This technique utilizes fluorescence to monitor the accumulation of the amplicon during the PCR and uses the fluorescent signal to determine the target quantity at the beginning of the reaction. During the initial few cycles, the fluorescent signal is too weak and cannot be distinguished from the background noise produced by the instrument. However, as the amount of PCR products accumulate exponentially, the signal strength increases proportionally. Quantitative data are generated using the cycle threshold, or C_t, value at which the fluorescent signal is detected and differentiated from background noise signals.

A real-time PCR reaction profile has three different parts: background phase, exponential (log) phase, and a plateau. The background phase is also known as the early phase, where the fluorescent signal from the amplicon is less than or equal to the background signal of the system. When there is sufficient accumulation of the PCR products, the reaction enters the exponential phase at which point the fluorescent signal exceeds the background signal. The reaction

efficiency eventually falls due to depletion of reagents and the reaction is in the plateau phase. In real-time PCR quantitation, the exponential phase is the most important and the amplification of the target is described by the following equation:

$$T_n = T_0(1+E)^n$$

In this equation, T_n refers to the amount of target amplicon at cycle number *n*, T_0 is the initial amount of target, and E is the efficiency coefficient of the amplification. The maximum efficiency achievable in PCR is 1, which means the target amplicon is doubling each cycle. The fluorescent signals given off by the probe are represented as curves in the real-time PCR instrument. The cycle number at which the signal goes above the background noise is dependent on the starting amount of template present at the beginning of the reaction. Therefore, the higher the initial concentration of the template, the earlier the fluorescent signal from the reporter dye will overcome the background fluorescence, indicated by an earlier C_t value. This C_t value is an arbitrary value and it is completely dependent on the threshold line. Due to this arbitrary nature, it is often difficult to compare real-time PCR results between different runs. The resulting C_t values can be used to develop a standard curve for absolute quantification or to determine fold differences for relative quantification.

In absolute quantitation, a DNA template of a known concentration is diluted five to seven times, usually in 2-fold or 10-fold increments, to generate a standard curve that relates the initial concentration of the template to the C_t value. The C_t value resulting from templates of unknown concentrations can then be used to estimate the absolute copy numbers of the unknown sample. In relative quantitation, a standard curve is generally unneeded because the absolute copy number is not the interest. Rather, this method calculates the fold difference between various samples to a single “reference” sample. Relative quantitation is most often used in drug

research, where the gene expression may be measured in response to a drug, which is compared to an untreated control group.

2.3.2 Non-specific DNA Dyes: SYBR Green I Chemistry

One of the most popular fluorescence chemistry used in real-time PCR is the SYBR Green chemistry (Morrison, Weis, & Wittwer, 1998), a type of non-sequence-specific DNA dye. SYBR Green is an asymmetric cyanine dye (Zipper, Brunner, Bernhagen, & Vitzthum, 2004), which consist of two nitrogen-containing aromatic compounds: one of which is positively charged (Figure 4). This type of dye emits minute amounts of fluorescence in their free state and when the dye binds to the minor groove of a ds-DNA, the rotation between the aromatic components are inhibited, and this causes the emission of a strong fluorescence (Nygren, Svanvik, & Kubista, 1998). One of the advantages of using the SYBR Green I system is that the chemistry is accessible for a moderate price, relative to other real-time PCR chemistries. Another advantage is that this chemistry gives access to a melting curve analysis, a process in which the melting temperature, or T_m, of the PCR products are calculated. This analysis gives some information on the identity of the products, which is useful in determining the specificity of the reaction. Information on the identity of the PCR products becomes important for this chemistry because SYBR Green I intercalates with any ds-DNA. Therefore, users must be wary of non-specific binding, including primer dimers and falsely-primed complexes. There is also the possibility of multiple dye molecules binding to the same DNA molecule simultaneously, meaning that the mass of DNA (number of base pairs) can influence the fluorescence measured by the instrument. Therefore, although SYBR Green I is a cheap and fast option, the quantitative results may not be accurate. Users must design the primers carefully to minimize non-specific

binding and perform a melting curve analysis to ensure that the PCR products are the intended target.

2.3.3 Sequence-specific hydrolysis probes: TaqMan Chemistry

The TaqMan probe is a hydrolysis probe that requires for the probe to be hydrolyzed by the 5' to 3' exonuclease activity of the DNA polymerase for activation (Heid, Stevens, Livak, & Williams, 1996). Therefore, the polymerase used for this chemistry must have exonuclease activity. The *Taq* polymerase is one example of a polymerase with this characteristic. This chemistry requires two sequence-specific primers, much like other PCR (and real-time PCR) systems, and a sequence-specific probe, adding an extra dimension of specificity to the reaction. The probe is dual-labeled, containing a fluorescent reporter dye at the 5' end (e.g. FAM, HEX, and TET) and a second dye, called the quencher, at the 3' end (e.g. TAMRA and IBFQ). The function of the quencher is to absorb the fluorescence emission from the fluorescent reporter dye, via fluorescence resonance energy transfer (FRET), when the two dyes are at proximity. Recently, probes that are double-quenched, with two labeled quenchers, have been developed (e.g. ZEN). While traditional probes have approximately 20-30 bases between the two dyes, an addition of a third internal quencher decreases this length to about 10 bases. This, combined with the original quencher, leads to increased quenching activity and thus reduces the background fluorescence, while increasing both sensitivity and precision of the experiment. An internal quencher also allows for longer probes to be designed.

The basic TaqMan chemistry starts with the probe, followed by the primers, hybridizing to the target region of the single-stranded template DNA (Figure 5). During the elongation phase, the polymerase moves along the template, adding dNTPs to the primers. When the

polymerase reaches the probe on the template, it hydrolyzes the probe, causing the release of the reporter from the quencher. The fluorescence emission from the reporter dye is no longer absorbed by the quencher and the accumulation of the un-quenched reporters lead to a strong fluorescent signal, which is detected by the real-time PCR instrument (Heid et al., 1996; Lyamichev & Dahlberg, 1993).

2.3.4 Designing primers and probes for real-time PCR

In real-time PCR, it is recommended to limit the size of the resulting amplicon to 75-200 base pairs. Smaller amplicons have lower melting temperatures (T_m), allowing for the DNA strands to be separated within shorter time. Having smaller amplicons also increase the efficiency of hybridization between the template and the oligonucleotides (primers and probes). Primers designed for the real-time PCR method are usually about 15 to 25 base pairs in length, have about 50-60% GC content, and have T_m of 57-61°C. It is also recommended for the primer pairs to not have more than 1-2°C difference in T_m. Real-time PCR probes are usually about 30 base pairs in length and designed to have a T_m that is 6-8°C higher than its corresponding primer pairs. This higher T_m is important because it ensures the probe binding to the target prior to the primers (as the temperature decreases from the denaturation to the annealing step) and competing efficiently against the more abundant primers.

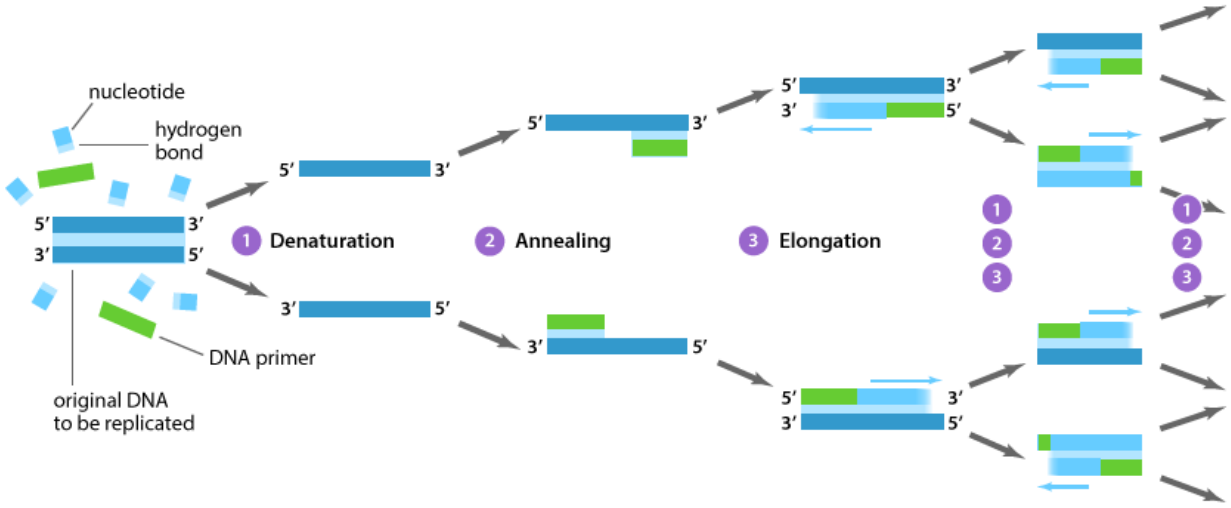

Figure 3. Thermal cycling steps of PCR. PCR has three thermal steps: denaturation, annealing, and elongation (extension). The PCR process usually repeats for 35 to 45 cycles. Image obtained from ABM (n.d.).

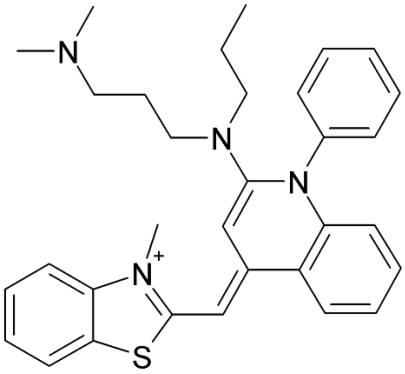

Figure 4. Chemical structure of SYBR Green I.

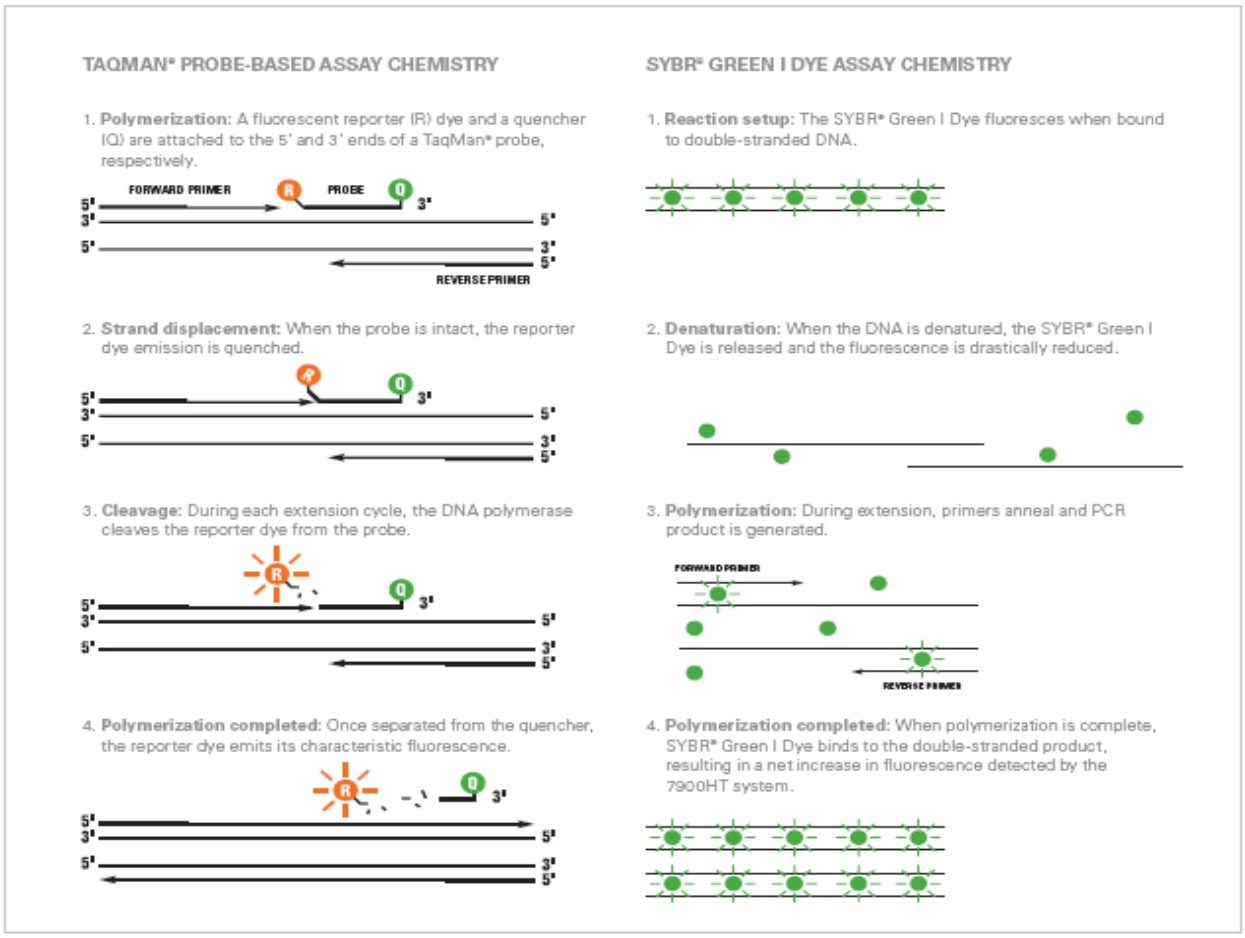

Figure 5. Differences between the TaqMan and SYBR Green chemistries. This figure was obtained from the Bio-synthesis, Inc. website (Bio-Synthesis, n.d.)

CHAPTER 3. DESIGNING PRIMERS AND PROBES FOR THE *PAPAIN*, *COAT PROTEIN*, AND MALE-SPECIFIC MARKER

3.1 Concept

The goal of this thesis project is to develop a methodology that can distinguish seeds pollinated by hermaphrodite plants from those pollinated by male plants. Thus, it is necessary to establish a genetic marker specific for each of the pollen parent genomes, then design an appropriate assay to detect and quantify signals from these markers. Quantitative PCR is a fast and high throughput method for accurately quantifying specific targets within a genome (Bubner & Baldwin, 2004; Ginzinger, 2002). To utilize this technology, sequence-specific primers and fluorescent probes that bind to unique sequences differentiating the hermaphrodite and male pollen signals need to be designed. As discussed in Chapter 1 (Figure 2), the hermaphrodite pollen contribution in the bulked seeds will be tracked using the *coat protein* gene, which is a component of the transgenic construct in GE papaya. The male pollen contribution will be tracked using a DNA marker sequence unique to the male papaya genome. In addition, the *papain* gene, an endopeptidase enzyme specific to the papaya species (Xu et al., 2008), will be used as an endogenous control for normalizing variation in initial DNA concentration in qPCR assays.

3.2 *Papain* and *coat protein* oligonucleotides

The *papain* gene has been demonstrated to be a reliable endogenous reference gene for papaya (Xu et al., 2008; Zhu et al., 2012). The primer and probe sequences for the *papain* and *coat protein* genes have been previously designed for the Chinese transgenic papaya ‘Zhonkang’

using the ABI Prism Primer Express Version 2.0 software (Applied Bio-systems, Foster City, USA) (Nageswara-Rao et al., 2013; Xu et al., 2008) and the sequences are shown in Table 3. The *papain* primers amplify a 91-base pair fragment and the *coat protein* primers amplify a 100-base pair fragment.

Table 3. qPCR primers and probes for the *coat protein* and *papain* genes designed by Xu *et al*. (2008) from sequence in papaya cultivar 'Zhongkang'.

Primer/ Probe	**Sequence (5' to 3')**
cp forward	CCG CGG TAT GGA ATC AAG AG
cp reverse	TCG AGA GCC ATA TCA GGT GTT TT
cp probe	(FAM)-CTC GCT AGA TAT GCT TTC GAT TTC TAT GCG GT-(MGB)
papain forward	AGT GGC TCA ATA TGG TAT TCA CTA CAG A
papain reverse	AAA ATG TAG ATA TAC CTC CCT TGA GCG
papain probe	(FAM)-ATA CTT ACC CAT ATG AGG GAG TGC AAC GTT ATT G-(TAMRA)

Since the two sets of primers and probes were designed for a Chinese cultivar, it was necessary to verify that the oligonucleotides function properly for Hawaiian cultivars, which represent a different genetic lineage and are genetically engineered with a different *coat protein* construct. The *papain* primers had been previously tested using genomic DNA template from Hawaiian cultivars, and efficient amplification resulted (Nageswara-Rao et al., 2013). Appendix A has the sequence of the *papain* gene with the primer and probe locations highlighted. The *coat*

protein primers had also been used successfully with Hawaiian papayas (Nageswara-Rao et al., 2013). However, a BLAST search of PRSV *coat protein* constructs showed small differences in the sequences employed in transforming 'Zhonkang' and 'Rainbow'. In both cases, the published reverse primer and probe sequences (Nageswara-Rao et al., 2013; Xu et al., 2008) had 2 to 3 nucleotide differences from what was expected for the Hawaiian papaya. Consequently, new primer pairs and probes were designed, using the 'Rainbow' papaya (Hawaiian cultivar) sequence available on GenBank (GenBank No.: FJ467933.1). The relevant portion of the sequence is shown in Appendix A. The sequences of the *papain* and *coat protein* primer and probes that were ordered for this project are shown in Table 4.

Table 4. qPCR primer and probe sequences designed for the *coat protein* and *papain* gene used in this thesis. The *coat protein* amplicon is 100 bp and the *papain* amplicon is 91 bp. The probes are designed with double quenching technology (ZEN as an internal quencher) from IDT (IDT, Coralville, IA).

Primer/ Probe	**Sequence (5' to 3')**
cp forward	CCG CGG TAT GGA ATC AAG AG
cp reverse	TCG CGA GCC CTA TCA GGT GTT TT
cp probe	(5' 6-FAM) – CTC GCT AGA (ZEN) TAC GCT TTC GAC TTC TAT GAG GT – (3' Iowa Black FQ)
papain forward	AGT GGC TCA ATA TGG TAT TCA CTA CAG A
papain reverse	AAA ATG TAG ATA TAC CTC CCT TGA GCG
papain probe	(5' HEX) - ATA CTT ACC (ZEN) CAT ATG AGG GAG TGC AAC GTT ATT G - (3' Iowa Black FQ)

3.3 Designing primers and a probe for a male-specific marker

3.3.1 Introduction

Identification of sex-specific markers for the papaya species is a very current research topic. In a recent publication by Liao et al. (2017), two regions unique to the male papaya genome were identified and, named PMSM-1 (Papaya Male-Specific Marker-1) and PMSM-2. Primer pairs that amplify male-specific sequences within these regions were developed, yielding a 585-bp and 548-bp PCR amplicons, respectively (Table 5). Unfortunately, these products are

too large for accurate qPCR analysis (Thornton & Basu, 2011). Therefore, new primer pairs needed to be designed to meet qPCR specifications. In addition, a probe that anneals to the sequence between the two priming sites needed to be designed.

Table 5. Standard PCR primers that amplify a 585-bp and 548-bp fragment from PMSM-1 and PMSM-2, respectively (Liao et al., 2017).

Primer	**Sequence (5' to 3')**
PMSM-1 forward	GGT CCA TAA GAC TCC TGA AG
PMSM-1 reverse	TCG TGT TGT TAG GCC AAG TG
PMSM-2 forward	GCG ATG CTT CAA GTG TTG AC
PMSM-2 reverse	ACT ATG AGC CTC ACG CAC TA

3.3.2 Materials and Methods

3.3.2.1 Primer Design

The 585-bp amplicon from PMSM-1 was searched using a BLAST search to confirm its unique origin. The result gave a single complete BAC clone sequence (96,769 bp) of the *Carica papaya* Y chromosome (GenBank No.: EU369761.1). Sequences of the 585-bp and 548-bp amplicons originating from PMSM-1 and PMSM-2 were found within this entry. The 585-bp sequence for PMSM-1 was used to generate new primer and probe candidates, using the *Primer 3 Plus* software with standard qPCR primer design criteria (Table 6). The target amplicon size in a qPCR assay needs to be much smaller than that of standard PCR and is optimally around 80-

150 base pairs (Thornton & Basu, 2011). The primer size is optimally about 20 base pairs and the probe is about 30 base pairs (Thornton & Basu, 2011). The difference in the size of the oligonucleotides causes a difference in their T_m's. Probes are typically designed to have a 6-8°C higher T_m than the primers so that the probes will anneal to the template DNA before the primers. This will ensure that the probes are perfectly hybridized to the template before the DNA polymerase reaches the probe during the amplicon extension phase. The GC%, CG clamp, and max end GC are characteristics of the oligonucleotides that deal with the guanine and cytosine nucleotides. Typically, it is recommended that about 50% of the oligonucleotide consist of guanine or cytosine residues (Thornton & Basu, 2011), because when the GC% is low, the T_m of the primers becomes too low. As for the max end GC, this refers to the number of guanine or cytosine residues to allow within the last 5 base pairs of the primers. Guanine and cytosine residues have a stronger bonding energy when they anneal; therefore, it is recommended that 2 or 3 of the last 5 residues be C or G ("Primer Design Guide for PCR :: Learn Designing Primers for PCR," ; Thornton & Basu, 2011).

The searches for primer and probe candidates were conducted on two different dates: 2/23/2016 and 3/9/2016. The specifications for the latter search were different from those shown in Table 6, in that the primer T_m was increased to 60 to 65°C with an optimum of 62°C and the probe T_m was increased to 65 to 75°C with an optimum of 70°C. This secondary search was conducted to increase the number of candidates.

Table 6. Design criteria for qPCR primers and probes, using *Primer 3 Plus* software.

Product Size: 80 to 150 bp	**CG Clamp: 1**
Primer Size: 18 to 23; optimum 20 bp	**Max End GC: 3**
Primer T_m: 57 to 63; optimum 60°C	**Probe Size: 27 to 35; optimum 30 bp**
Max T_m Difference: 2°C	**Probe T_m: 62 to 72, optimum 68°C**
Primer GC%: 30-80; optimum 50%	**Probe GC%: 30-80, optimum 50%**

3.3.2.2 BLAST Search

The male-specificity of the primer (both forward and reverse) and probe candidates were verified using a BLAST search. If any of the candidates had high sequence identity with a non-male papaya genome, the candidates were discarded from the list. The remaining candidates were checked for presence of secondary structures (candidates with Gibbs free energy value less than -8 kcal/mol were discarded). Out of the remaining candidates, four primer pairs with the least homology to non-male papaya genome sequences were ordered for testing.

3.3.2.3 DNA Extraction

When the four primer pairs arrived, DNA was extracted from papaya leaf tissue using the method published in Xin, Velten, Oliver, and Burke (2003). Briefly, one freshly punched 0.5-cm diameter papaya leaf disc was incubated at 95°C for 10 minutes in a 200-µL microfuge tube containing 50 µL of buffer A (100 mM NaOH, 2% Tween 20). Following addition of 50 µL of pH-neutralizing buffer B [100 mM Tris-HCl, 2 mM EDTA], the tube was vortexed at moderate speed, and the contents were used as template DNA for PCR.

This method was used to extract DNA from six papaya leaves, including two papaya leaves for each sex. The female leaves came from "Line 34 x Sekaki" and "Sekaki x Kapoho". The hermaphrodite leaves came from "Sekaki x Kapoho" and "Big Island x Kapoho". Both male leaves were collected from a single feral plant located near the Waimanalo Experiment Station (O'ahu, Hawai'i).

3.3.2.4 PCR specifications

Four primer pairs used in this section were synthesized by Integrated DNA Technologies (IDT, Coralville, IA, USA). The primers arrived lyophilized with approximately 25 nanomoles of DNA. Each primer was diluted with ultra-pure water to produce a 100 µM stock solution, which was then diluted further to obtain several tubes with 10 µM primers. PCR amplification was carried out using a programmable thermal cycler (Applied Biosystems GeneAmp PCR System 9700, Applied Biosystems, Waltham, MA, USA) in 25-µL reaction mixture volumes containing 12.5 µL of 2x GoTaq Green master mix (Promega, Madison, WI, USA), 1.0 µL each of the forward and reverse primers (final concentration of 400 nM), 9.5 µL of water, and 1 µL of the template. The reactions were run at 95°C for 5 minutes (initial denaturation) followed by 35 cycles of 95°C for 15 seconds (denaturation), 57°C for 15 seconds (annealing), and 72°C for 45 seconds (extension). The final extension phase was set at 72°C for 4 minutes.

The number of reactions were 28, with 6 different templates (3 sexes, 2 samples per sex), 4 different master mixes (1 for each primer pair tested), and 4 water control samples (one for each master mix).

3.3.2.5 Gel Electrophoresis

The amplified PCR products were resolved on a 1.5% agarose gel with 1x TBE (Tris-HCl, Boric acid, EDTA) at pH 8.3. Twenty-five wells were used, including 24 samples and one 100-bp Hyper Ladder IV (Bioline, London, UK), with 5 µL of the PCR product or ladder loaded into each well. The gel was run for 45 minutes at 60 V. The gel was stained in ethidium bromide (0.5 µg/mL) for 10 minutes and de-stained in water for 15 minutes. The gel image was viewed and recorded under UV light.

3.3.2.6 Genomic Sequencing

The PCR amplicons were treated with the USB ExoSAP-IT PCR Product Cleanup kit (Affymetrix, Santa Clara, CA, USA) according to the manufacturer's specifications, and the purified products with forward and reverse primers (1.6 µM) were sent to the University of Hawai'i Advanced Studies in Genomics, Proteomics, and Bioinformatics (ASGPB) for sequencing.

3.3.2.7 Analysis of Sequencing Data

The sequencing results from the ASGPB were viewed using Chromas Lite (Technelysium Pty Ltd, version 2.5.0) and the resulting nucleotide sequences were aligned using the Molecular Evolutionary Genetics Analysis (MEGA) software (version 6.0.0).

3.3.3 Results

3.3.3.1 *Primer 3 Plus* results

The results of the two *Primer 3 Plus* searches based on different T_m criteria are shown in Appendix B. Candidate primers identified with the specifications in Table 6 were labeled "low temp" and those identified with the higher temperature specification were labeled "high temp".

The 18 candidate primer pairs were tested for specificity to the male genome and for any presence of potential secondary structures. Four of the best candidates based on male genome specificity and low potential for secondary structures were Pair 1 (Low Temp), Pair 6 (High Temp), Pair 7 (High Temp), and Pair 8 (High Temp). They were renamed to Pair 1 (originally pair 1 Low temp), Pair 2 (originally pair 6 high temp), Pair 3 (originally pair 7 high temp), and Pair 4 (originally pair 8 high temp). Figure 6 shows the location of the priming and probe sites for each candidate primer pair within the 585-bp amplicon of PMSM-1.

Pair 1

ATGGCCTATCACGTATGGGTTCCCCACCTGCATAGCCAAGCATACAACACAAACGA
GTAACCAAGCCCACACGGTCCATGGCATGCCTACAAGCATTCGAGAAAGCACTTAC
ACATTCAGCCCATGCACATGTGTGGTCAAGGCATTCTACACACCCTGCGTGCATAC
GAAAATGCGCAGGAGAAGCCCCACATATATGCGACAATAGGCATACTCAATCCACT
CCTAAAAGACTACCCAATAACCCACTTGGCCTAACAACACGA

Pair 2

ATGGCCTATCACGTATGGGTTCCCCACCTGCATAGCCAAGCATACAACACAAACGA
GTAACCAAGCCCACACGGTCCATGGCATGCCTACAAGCATTCGAGAAAGCACTTAC
ACATTCAGCCCATGCACATGTGTGGTCAAGGCATTCTACACACCCTGCGTGCATAC
GAAAATGCGCAGGAGAAGCCCCACATATATGCGACAATAGGCATACTCAATCCACT
CCTAAAAGACTACCCAATAACCCACTTGGCCTAACAACACGA

Pair 3

ATGGCCTATCACGTATGGGTTCCCCACCTGCATAGCCAAGCATACAACACAAACGA
GTAACCAAGCCCACACGGTCCATGGCATGCCTACAAGCATTCGAGAAAGCACTTAC
ACATTCAGCCCATGCACATGTGTGGTCAAGGCATTCTACACACCCTGCGTGCATAC
GAAAATGCGCAGGAGAAGCCCCACATATATGCGACAATAGGCATACTCAATCCACT
CCTAAAAGACTACCCAATAACCCACTTGGCCTAACAACACGA

Pair 4

ATGGCCTATCACGTATGGGTTCCCCACCTGCATAGCCAAGCATACAACACAAACGA
GTAACCAAGCCCACACGGTCCATGGCATGCCTACAAGCATTCGAGAAAGCACTTAC
ACATTCAGCCCATGCACATGTGTGGTCAAGGCATTCTACACACCCTGCGTGCATAC
GAAAATGCGCAGGAGAAGCCCCACATATATGCGACAATAGGCATACTCAATCCACT
CCTAAAAGACTACCCAATAACCCACTTGGCCTAACAACACGA

Figure 6. Annealing sites of the four primer/probe candidates within the 585-bp amplicon from PMSM-1. Forward primer is in yellow, probe is in green, and the reverse primer is in blue.

3.3.3.2 Verification of male-specificity by PCR

To verify the male-specificity of the primer pairs that were developed, leaf DNA was extracted from each of the three papaya sexes (two different sources for each sex) and the target sequence was amplified by standard PCR, using the four candidate primer pairs. The PCR products were resolved on a 1.5% agarose gel viewed under UV light (Figure 7). All four primer pairs produced a strong product approximately 100-bp in length when amplifying male DNA template. The relative positions of the bands were consistent with the expected amplicon size for each of the primer pairs (Pair 1: 148 base pairs, Pair 2: 88 base pairs, Pair 3: 118 base pairs, Pair 4: 106 base pairs). There were faint bands present in the female and hermaphrodite template lanes of pairs 2 and 4, indicating some undesirable amplification of non-male genomes by these primer pairs. There was also a thin 750-bp band in one of the female templates amplified by pair 3, indicating potential for non-male-specific priming. Pair 1 appeared to be the best candidate out of the four pairs, based on the intensity of the banding of the male marker and on its specificity to the male papaya genome.

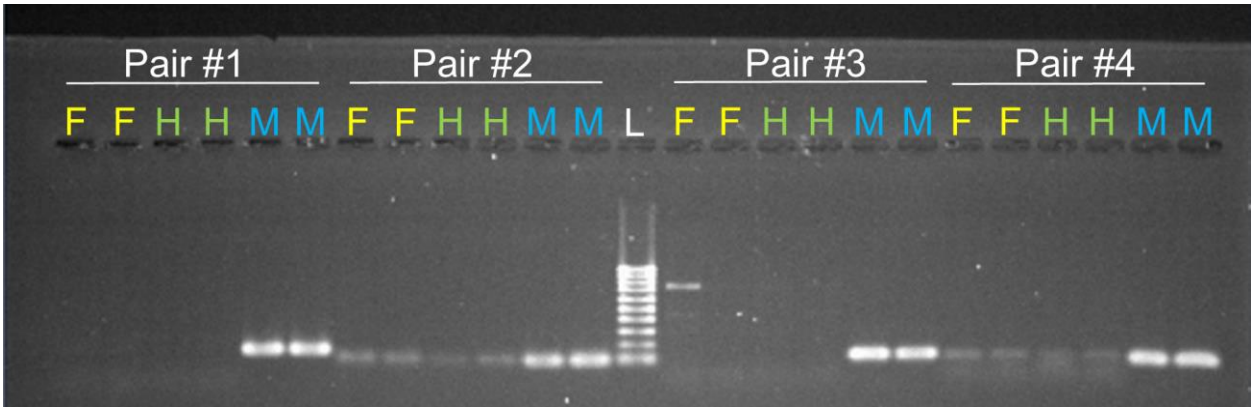

Figure 7. PCR products resolved on a 1.5% agarose gel viewed under UV light. Pair number refers to which male-specific primers were used. The letters under the pair number indicate the plant sex from which the DNA template was extracted (F=female, H=hermaphrodites, M=males).

3.3.3.3 Verification by Sequencing

All PCR products amplified from the male leaf DNA template were sent to the University of Hawai'i at Mānoa ASGPB. The electronic data were analyzed on Chromas Lite (version 2.5.0) and aligned using MEGA (version 6.0.0). Figure 8 shows the aligned sequences. A 100% match between the expected probe sequence and the entries was found in two instances, once in the amplicon from pair 1 and once in the amplicon from pair 3.

Due to technical difficulties and limitations during the sequencing process, about 10 to 15 base pairs on both ends of the sequence were lost. As a result, the expected probe sequence could not be verified in many of the amplicons sent in for testing.

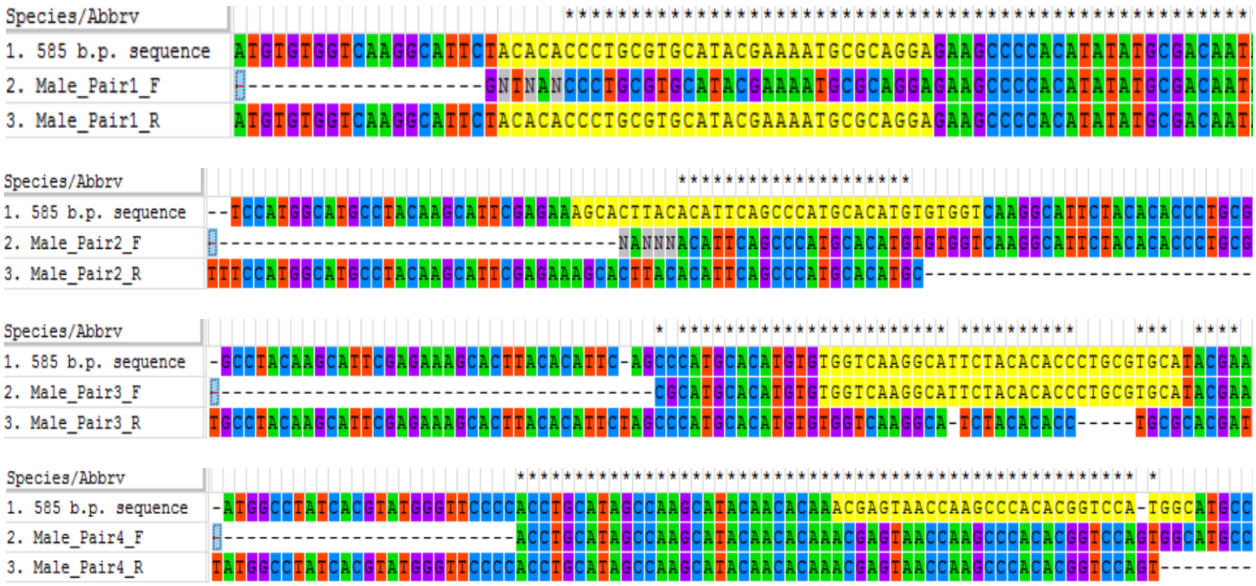

Figure 8. Alignment of the sequencing results of male PCR products using MEGA. The first line for each entry is the expected sequence (585 base pairs) within PMSM-1. The second line for each entry is the sequence result with the forward primer. The third line for each entry is the sequence result with the reverse primer. The probe site is highlighted in yellow.

3.3.4 Conclusion

In conclusion, three verification methods including BLAST search, agarose gel visualization, and Sanger sequencing were utilized to test the primer candidates for specificity to the male papaya genome. After these tests, it was determined that Pair 1 (low temp) was the best candidate to use for the qPCR analysis. The 148-bp sequence that is amplified by this primer pair will be referred to as the *male* marker in the coming chapters.

CHAPTER 4. DNA EXTRACTION FROM PAPAYA TISSUE

4.1 Introduction

In the past few decades, researchers developed numerous methods for extracting high quality DNA from plant tissues for molecular analysis (Doyle & Doyle, 1987; Murray & Thompson, 1980). These researchers were looking to efficiently remove impurities such as polysaccharides, tannins, and proteins, which were difficult to separate from DNA (Murray & Thompson, 1980), while protecting the DNA from nucleases and mechanical shearing (Varma, Padh, & Shrivastava, 2007). Many of the protocols used today use buffers containing CTAB, or cetyl-trimethyl-ammonium-bromide, which is a surfactant with antiseptic properties. CTAB buffers are able to effectively remove lipids of the cell membrane and promote cell lysis, and can also bind to polysaccharides and remove them from solution (Clarke, 2009). However, there is no universal DNA extraction method that works equally well for all plants.

An extensive literature search revealed three manuscripts that described methodologies for extracting high quality DNA from papaya seeds that can be used for qPCR analysis (Guo et al., 2009; Nageswara-Rao et al., 2013; Xu et al., 2008). Xu et al. (2008) and Nageswara-Rao et al. (2013) utilized the CTAB method, while Guo et al. (2009) utilized a commercially available mini-plant genomic DNA extraction kit (Shanghai Ruifeng Agro-tech Co. Ltd, Shanghai, China).

4.2 Materials and Methods

4.2.1 Papaya Seed and Leaf Sources

As described in Chapter 2, the genetic control of sex expression in papaya in papaya dictates that the males (X/Y^m) and hermaphrodites (X/Y^h) are heterozygous, and therefore they cannot produce seed progenies that are uniformly male or hermaphrodite by self- or cross-pollination [homozygous Y genotypes (Y^m/Y^m, Y^h/Y^h, or Y^m/Y^h) are lethal and do not exist]. Consequently, seed batches always segregate for approximate ratios of at least two sexes, so they are not useful as sources of pure DNA of any single sex for development of marker standards for qPCR.

In this thesis project, the transgenic *cp* gene was used as a proxy marker for the hermaphrodite sex, although there is no genetic linkage between sex and the transgene. The *cp* gene is unique to transgenic lines and functions well as a marker for pollen movement in a matrix of non-transgenic receptor plants. Unlike the sex genes, *cp* can be produced in seed batches in uniform genetic dosage, which is necessary for obtaining pure marker DNA standards for qPCR.

Seeds for extracting DNA were obtained from fruits of gynodioecious papaya lines harvested in 2013 from Waimanalo Research Station on O'ahu. Seeds had been cleaned (sarcotestas removed), air-dried, and stored at 4°C. All seeds were F_1 hybrid, produced by crossing a non-transgenic parent with a homozygous (*cp*/*cp*) transgenic parent, and therefore the resulting F_1 seeds were uniformly hemizygous (*cp*/-) for the *cp* gene. Various non-transgenic parents were used in the crosses, including 'Sekaki' and 'Big Island', and the transgenic parent was 'Line 34', a homozygous transgenic (*cp*/*cp*) breeding line. All parental lines were

gynodioecious, and all hybridized female flowers were bagged at the time of crossing to prevent contamination from unwanted pollen sources.

Since the marker for male sex is linked to the male-specific region of the Y^m genome, seed batches pollinated by male trees will segregate for the marker in approximate ratios unsuitable for developing DNA standards for qPCR. Instead, DNA was extracted from young leaves of feral male plants collected in Waimanalo, O'ahu. The somatic tissue is uniformly heterozygous for the marker sequence (X/Y^m), so it has the same genetic dosage for the male as the hemizygous *coat protein* (*cp*/-) gene in F_1 seed has for the hermaphrodite.

DNA was also extracted from leaves of transgenic 'Rainbow' papaya plants, which are hemizygous (*cp*/-) for the transgene. DNA standards developed from this source were used to check for differences in seed and leaf standard curves attributable to tissue type.

4.2.2 DNA Extraction from Papaya Leaf

DNA extraction from papaya leaf was performed using the DNeasy Plant Mini Kit (Qiagen Inc., Valencia, CA, USA). Young leaf tissue is usually a preferred maturation stage for DNA extraction because there is less deposition of starch and secondary metabolites (Peterson, Boehm, & Stack, 1997; Williams & Ronald, 1994). In contrast, DNA extracted from mature leaves is reported to be of poor quality due to high amounts of polysaccharides, tannins, and other metabolites (Hanania, Velcheva, Sahar, & Perl, 2004; Sangwan, Sangwan, & Kumar, 1998; Sharma, Lavanya, & Anjaiah, 2000). Approximately 150 mg of young papaya leaves were used for this method. The procedure was conducted according to manufacturer's instructions.

4.2.3 DNA Extraction from Papaya Seeds

Papaya seeds, unlike leaf tissue, are a difficult source from which to extract DNA for PCR use because there are many impurities that interfere with biological enzymes (Fang, Hammar, & Grumet, 1992; Tel-zur, Abbo, Myslabodski, & Mizrahi, 1999; Varma et al., 2007). CTAB extractions are the most widely used method for papaya seeds because it removes most of these impurities (Nageswara-Rao et al., 2013; Xu et al., 2008). There are many versions of this method, and this thesis sought to optimize this method to get high and consistent DNA yields.

The first CTAB extraction procedure tested for this thesis came directly from the manuscript by Nageswara-Rao et al. (2013). The full procedure can be found in Appendix D. Briefly, about 5 grams of papaya seeds were macerated using a mortar and pestle under liquid nitrogen. The seed powder was added to hot CTAB buffer and incubated at 65°C for 45 minutes, with intermittent shaking. The solution was spun down to separate the debris, and 1 mL of the supernatant was emulsified with an equal volume of phenol: chloroform: isoamyl alcohol (25:24:1) twice. This is followed by another emulsification with chloroform: isoamyl alcohol (24:1) twice. Genomic DNA was then precipitated using two volumes of ice-cold isopropanol, and the pellet was washed with 70% ethanol twice. Then, the pellet was suspended in 100 µL TE buffer. The DNA solution was treated with 4 µL of RNase A (10 mg/mL) and precipitated once again with two volumes of isopropanol and one-tenth volume of 7.5 M sodium acetate. The pellet was washed with 100 µL of 70% ethanol and re-suspended in 50 µL of TE buffer.

Another protocol, developed by the Pacific Basin Agricultural Research Center (Hilo, Hawai'i), utilizes a combination of the CTAB buffer method and the DNeasy kit. The procedure is briefly described in the manuscript by Matsumoto, Zee, Suzuki, Tripathi, and Carr (2010). The procedure in this thesis was adapted from this protocol and optimized for the needs of this

project (Appendix E). Briefly, 500 seeds (approximately 8 grams) were macerated with a mortar and pestle under liquid nitrogen. The seed powder was transferred to 100 mL of hot CTAB buffer and the mixture was incubated for 30 minutes, with agitation every 5 minutes. Seed residues were removed by filtering through Miracloth, and 30 mL of the supernatant was emulsified with 20 mL of chloroform: isoamyl alcohol (24:1). The aqueous layer was precipitated with 2/3 volume ice-cold isopropanol and 2.5 mL of 3 M sodium acetate. The pellet was washed with 2 mL of 70% ethanol and re-suspended in 400 µL of TE buffer. The DNA was then further purified through the DNeasy Plant Mini Kit, following the manufacturer's instructions. Approximately 300 µL of DNA is eluted in the final step.

4.3 Results

Over the course of the thesis project, each method was tested numerous times to determine which method was the most reliable. DNA extraction from leaf tissue using the DNeasy kit, which usually took only 1 to 1.5 hours, was consistently able to yield DNA of approximately 30 ng/µL, with a 260/280 absorbance ratio between 1.67 and 1.92 (1.8 is optimal) from small amounts of leaf tissue (Table 4.1). DNA extraction from seed tissue using the 'Rao' protocol took approximately 7 hours on average and required multiple phenol/chloroform extractions. The average DNA yield calculated from about 20 different extraction attempts was approximately 90 ng/µL (Table 7). However, the yields were very inconsistent, ranging from as small as 10 ng/µL to as high as 500 ng/µL. There was a problem with purity, which also had high variation. DNA extraction from seeds using the 'Matsumoto' protocol took approximately 5 hours on average and did not require multiple extraction steps with expensive chemicals or hazardous phenol. The average DNA yield from about 30 different extraction attempts was 70

ng/µL (Table 7). The yield, on average, was much smaller than that of DNA extracted by the 'Rao' protocol, but the yields were much more consistent (Table 7). In fact, although the range of DNA concentrations recovered by the 'Matsumoto' protocol started at 14.4 ng/µL, this was only one extraction out of a total of 30, and the next lowest yield was 29.9 ng/µL. In addition, the range of DNA purity readings were narrower and more consistent from the 'Matsumoto' extracted seeds.

Above, two parameters that are commonly used to assess the quality, in terms of yield and purity, of DNA are discussed, but ultimately the most important factor is the efficacy of the extracted templates in PCR. Several template extractions prepared by each methodology were amplified with standard PCR using *papain* primers. The PCR products were resolved in a 1.5% agarose gel (Figure 9). As seen on the gel, all extracted DNA using the DNeasy kit and the 'Matsumoto' protocol were successfully amplified using *papain* primers. However, the extracted DNA from seeds using the 'Rao' protocol did not always amplify the *papain* amplicon successfully. Based on these results, it was concluded that the 'Matsumoto' protocol can extract useful template DNA from seeds more quickly, more reliably, and with less variation in the yield, compared to the 'Rao' protocol.

Table 7. Comparison of the three DNA extraction procedures.

Procedure	Tissue	Tissue weight (g)	Average DNA quantity (ng/µL)	DNA quantity range (ng/µL)	$A_{260/280}$
Rao	Seed	≈5.5	90	9.6 – 519.6	0.68 – 1.97
Matsumoto	Seed	≈8.0	70	14.4 – 153.5	1.64 – 1.97
DNeasy	Leaf	0.1	30	7.9 – 52.7	1.67 – 1.92

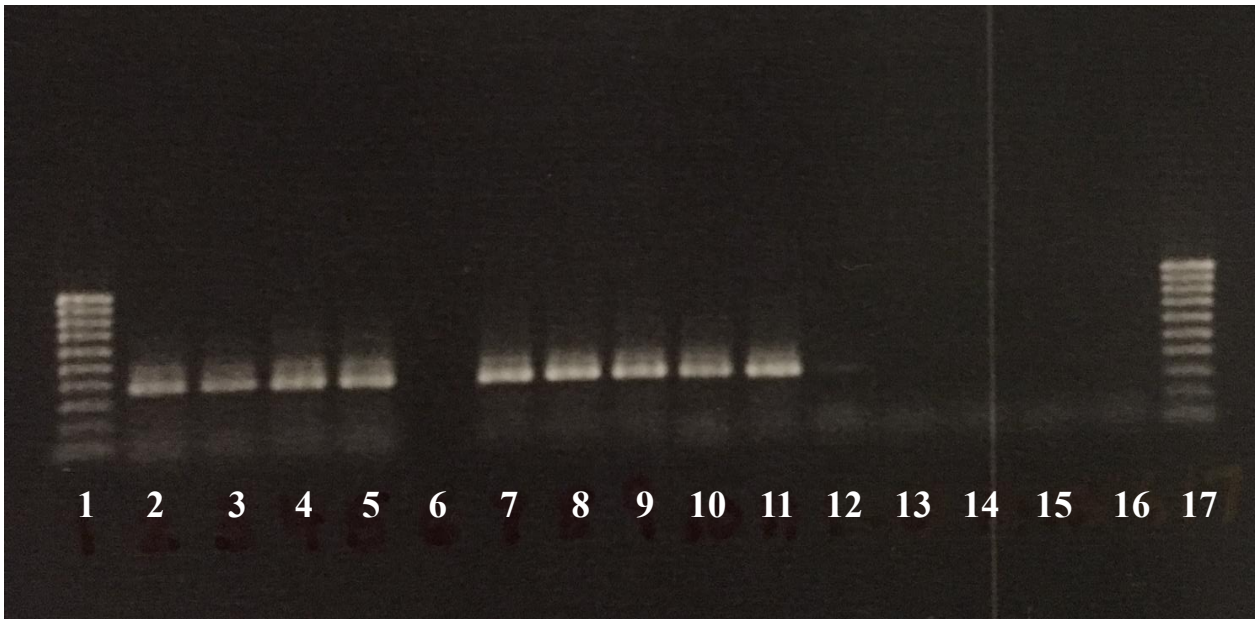

Figure 9. 1.5% agarose gel of PCR products with *papain* primers. Lane 1: 100-bp ladder, Lanes 2-5: hermaphrodite leaf DNA template (DNeasy; four different extractions), Lane 6: water control, Lanes 7-9: seed DNA extracted using the 'Matsumoto' protocol (three different extractions), Lanes 10-16: seed DNA extracted using the 'Rao' protocol (seven different extractions), Lane 17: 100-bp ladder.

CHAPTER 5. OPTIMIZATION OF THE QPCR STANDARD CURVE AND VALIDATION PROCESS

5.1 Introduction

Chapter 3 described the development of qPCR primers and probes that are specific for the target sequences (*cp* and *male*) and for an endogenous papaya reference gene (*papain*). All designed primers were successful in amplifying their respective targets in PCR. Chapter 4 outlined and discussed procedures for extracting high quality genomic DNA from papaya seeds and leaves containing the target gene sequences. The objective in Chapter 5 is to combine these elements in a qPCR assay that can distinguish and quantify the target gene DNA signals, and then interpret the data to determine what percentage of the total extracted DNA is attributable to pollen originating from hermaphrodite or male plants.

The qPCR combines the traditional PCR amplification of the target gene and the technology to monitor the amplification of the products in real-time. The amplification of the products can be quantified by monitoring fluorescent signals that bind to DNA. There are many different fluorophore binding and detection chemistries (Navarro, Serrano-Heras, Castaño, & Solera, 2015). The quantification method ultimately relates the magnitude of the fluorescent signal and the starting amount of target DNA in the sample. The fluorescent signal is detected by the instrument when it passes a certain threshold, and the number of amplification cycles it takes for each sample to reach this threshold (C_t) is related to the concentration of the DNA. For example, if two identical DNA templates were diluted to two different concentrations, the PCR with the higher concentration of DNA will reach this threshold earlier and have a lower C_t than the reaction with more dilute DNA.

Most qPCR analyses are performed by first constructing a standard curve for the target gene. The target gene at a known concentration is serially diluted several times, most commonly in 10-fold steps over a range of three or more orders of magnitude. Running the standards on a qPCR thermal cycler will allow the instrument to determine the C_t for each DNA standard and plot a standard curve, showing the relationship between the C_t values and the known concentrations of the target gene in each standard. By establishing this relationship, when a template with unknown concentrations of the target is tested, the instrument can use the resulting C_t value and calculate the concentration of the target. The standard curve must be validated by testing samples of known DNA composition.

The two target sequences in this study are the *coat protein* (*cp*) gene and the *male* marker. To create a qPCR assay to detect the *cp* gene, DNA from transgenic, hermaphrodite-pollinated seeds was used. Also, DNA from transgenic hermaphrodite leaves was used to examine the differences in tissue types. For the male assay, seeds could not be used for reasons discussed in Chapter 4. Instead, male leaf tissue was used as a substitute. After all the results were obtained, the relationship between the two types of tissue was determined and normalized.

5.2 Materials and Methods

5.2.1 Developing standard curves for the papain, coat protein, and male markers

5.2.1.1 Adjusting DNA concentration for the standard curve

To minimize variability in the concentration of total genomic DNA in the seed and leaf extracts from which each of the standard curves was produced, total DNA concentration was adjusted to 20 ng/µL using the Qubit 3.0 fluorometer (Thermofisher Scientific, Waltham, MA,

USA). The DNA concentration was first measured after extraction and again after diluting it to 20 ng/µL. Any deviation from this concentration was compensated for by adjusting the volume of DNA that was loaded into the 25 µL reaction volume. For example, if the target amount of DNA is 100 ng/reaction, 5 µL of DNA at 20 ng/µL was added to the reaction volume, but only 4.5 µL at 22 ng/µL.

The standard curve was made by performing a 10-fold dilution of the template DNA four times, so that the series consisted of 100 ng, 10 ng, 1 ng, 100 pg, and 10 pg of total genomic DNA per reaction. As mentioned previously, the volume of DNA loaded into the reaction tubes depended on the actual concentration of the DNA, as determined by Qubit fluorometry. The qPCR was performed following the protocols discussed below.

5.2.1.2 qPCR specifications

The optimization of a qPCR assay usually begins by using the SYBR Green I chemistry. This chemistry is relatively inexpensive (compared to probe-based chemistry) and allows the user to quickly assess the primer and DNA quality in the qPCR instrument. This chemistry was used in the thesis to test primers, but not to generate any data.

Gene-specific non-fluorescent primer pairs, along with fluorescent dye-labeled TaqMan probes were designed for the *papain*, *coat protein*, and *male* target sequences (Table 8). The PCR amplification was performed using the Rotor-Gene 6000 qPCR thermal cycler (Corbett Research, NSW, Australia) in a 25 µL reaction volume consisting of 12.5 µL of 2x Rotor-Gene Probe PCR Kit (includes buffer, $MgCl_2$, dNTPs), 1.0 µL each of 10 µM primers (final concentration of 400 nM), 0.5 µL of 10 µM probe (final concentration of 200 nM), 5 µL of ultra-

pure water and 5 µL of DNA template. The volume of DNA and water in the reaction volume was adjusted based on the concentration of the template.

The cycling conditions were as specified in the instructions for the Rotor-Gene Probe PCR Kit (Qiagen Inc., Valencia, CA, USA). The program consisted of 3 minutes at 95°C (initial denaturation & polymerase activation), followed by 55 cycles of 3 seconds at 95°C (denaturation) and 10 seconds at 60°C (annealing and extension). The results were viewed using the Rotor-Gene analysis program.

Table 8. Sequences of qPCR primers and probes used in this thesis.

Primer/ Probe	**Sequence (5' to 3')**
cp forward	CCG CGG TAT GGA ATC AAG AG
cp reverse	TCG CGA GCC CTA TCA GGT GTT TT
cp probe	(6-FAM) – CTC GCT AGA (ZEN) TAC GCT TTC GAC TTC TAT GAG GT – (Iowa Black FQ)
papain forward	AGT GGC TCA ATA TGG TAT TCA CTA CAG A
papain reverse	AAA ATG TAG ATA TAC CTC CCT TGA GCG
papain probe	(HEX) – ATA CTT ACC (ZEN) CAT ATG AGG GAG TGC AAC GTT ATT G – (Iowa Black FQ)
male forward	AGC CCA TGC ACA TGT GTG GTC AA
male reverse	TCG TGT TGT TAG CGG AAG TGG GT
male probe	(6-FAM) – ACA CAC CCT (ZEN) GCG TGC ATA CGA AAA TGC GCA GGA – (Iowa Black FQ)

5.2.1.3 Generating data using the Rotor-Gene program

At the completion of the qPCR cycling run, choosing the "analysis" option in the Rotor-Gene program enabled the instrument to find the optimal threshold level for the amplification results. Once the threshold level was determined, the program returned the C_t values from the amplification curves and plotted the standard curve relating the known concentrations of the template standards against their C_t values. Parameters of the standard curve, such as R^2, slope, y-intercept, and the efficiency of the reaction were viewable as well.

5.2.2 Validation

The assay developed in this thesis was validated by analyzing qPCR results of leaf DNA samples with known proportions of male and hermaphrodite targets. This experiment tested DNA mixtures in four different proportions, which were 1% hermaphrodite: 99% male, 10% hermaphrodite: 90% male, 90% hermaphrodite: 10% male, and 99% hermaphrodite: 1% male. The DNA's for both sexes were extracted from leaf tissues and adjusted to the same concentration before being mixed in the appropriate proportions. Then, the total DNA concentration for each of the four mixtures was measured and adjusted to 20 ng/µL. The qPCR reaction volumes and cycling parameters were identical to those used for the standard curve procedure. All experimental samples were assayed in triplicates.

5.2.3 Analyzing the data using the $2^{-\Delta\Delta Ct}$ method

The results from the qPCR were analyzed using the $2^{-\Delta\Delta Ct}$ method (Livak & Schmittgen, 2001), which calculates the target DNA amount in relative terms as a percentage of a known standard. The calculations are based on sample C_t values, which are first normalized to the

endogenous reference gene, *papain*, to adjust target signal strength for variation in total DNA amount from sample to sample. Subtracting the C_t value of the *papain* reference gene from the C_t for the target gene accomplishes the sample normalization by converting absolute C_t values to a relative difference, ΔC_t, that is independent of sample DNA concentration:

$$C_{t,\,target} - C_{t,\,reference} = \Delta C_t$$

The ΔC_t values are further adjusted to make them relative to the C_t of a normalized standard in which the target genome constitutes a known percentage of the total genomic DNA. In this project, the calibration standard was essentially 100% target genomic DNA, either *cp* or *male*, so the calibrated value, $\Delta\Delta C_t$, represents the target DNA amount of the normalized sample in terms of percentage of pure target DNA:

$$(C_{t,\,target} - C_{t,\,reference}) - \Delta C_{t,\,calibrator} = \Delta\Delta C_t$$

The $\Delta\Delta C_t$ can be converted to a "fold difference" in relative amplification of the target by plugging the value into $2^{-\Delta\Delta Ct}$, where the factor of 2 represents the assumed doubling of DNA per cycle of amplification. The $2^{-\Delta\Delta Ct}$ value is the fold difference in the target amplification, normalized to the *papain* reference gene and relative to a sample containing 100% target genome.

When this is applied to the calibrator ΔC_t, the $\Delta\Delta C_t$ value is equal to zero, yielding a fold difference of 1. Fold differences can be converted to percentages by dividing the fold difference of the calibrator by the fold difference of the sample and multiplying by 100:

$$100\ [(2^{-\Delta\Delta Ct})_{calibrator} / (2^{-\Delta\Delta Ct})_{sample}] = 100 / (2^{-\Delta\Delta Ct})_{sample} = \text{target gene \%}$$

In order for the $2^{-\Delta\Delta Ct}$ calculation to be valid, amplification efficiencies of the target and reference primers must be approximately equal (Livak & Schmittgen, 2001). To test this, the C_t of the target and reference genes is monitored over numerous dilution levels and the ΔC_t, which is the difference in C_t between the target and reference gene at the same dilution level, is calculated. After plotting the relationship (ΔC_t vs. DNA concentration), if the slopes from the three primer systems are similar and close to zero, the $\Delta\Delta C_t$ calculation for relative quantification of the target may be used. In this thesis, a 100-fold range, between 100 ng to 1 ng/reaction, is used for this analysis.

5.3 Results and Discussion

5.3.1 Amplification and standard curves

The amplification curves and the corresponding standard curves for hermaphrodite-pollinated seed (Figure 10), transgenic hermaphrodite leaf (Figure 11), and male leaf (Figure 12) tissues are shown below. The last dilution level (10 pg) for the male leaf did not show any amplification (Figure 12). The efficiency of the reaction for the hermaphrodite-pollinated seed, transgenic hermaphrodite leaf, and male leaf templates were 0.91, 0.62, and 0.75, respectively. The poor efficiency is likely due to high variation in the last dilution level on each graph. Efficiencies using only the first four dilution levels were 0.94, 0.70, and 0.75, respectively. Efficiencies using only the first three dilution levels were 0.95, 1.01, and 0.95. The assay showed a strong linear relationship in the ranges of 100 ng to 1 ng (100% to 1% of target) of total DNA, and although much lower concentrations could be detected, the efficiency of

amplification dropped dramatically. This suggests that the lower limit of accurate quantification on the standard curve is at about 1 ng of DNA. Practically, this would be equivalent to quantifying a single seed having the target marker in a batch of 100 non-target seeds. Since an average papaya produces about 500 seeds, this means that the assay may not be able to reliably quantify the presence of the target in fruits with fewer than about 5 seeds containing the target marker.

Quantitative PCR assumes doubling of the target DNA after each amplification cycle if the primer system is perfectly efficient, represented by the equation 2^n, where n represents the number of amplification cycles. The number of cycles it takes to see a 10-fold difference in the target gene concentration is approximately 3.3, because $2^{3.3} \approx 10$. The number of cycles between 10-fold dilutions on the standard curve (Table 9) ranged between 2.93 and 3.54 when considering the cycle differences between the first and second dilutions, as well as between the second and third dilutions. The cycle differences were much closer to the optimal value in data generated from leaf tissue (ranged between 3.07 to 3.54) compared to seed tissue (ranged between 2.93 to 3.27), in which the differences were lower than expected. The cycle differences between the third and fourth dilutions were inconsistent, ranging between 4.34 and 6.19, where all differences were higher than the expected 3.3 cycle difference. This pattern, in which the consistency of the cycle differences between dilution levels disappears after the third dilution level, is consistent with the efficiency of the reaction where the inclusion of the fourth dilution lowers the calculated efficiency.

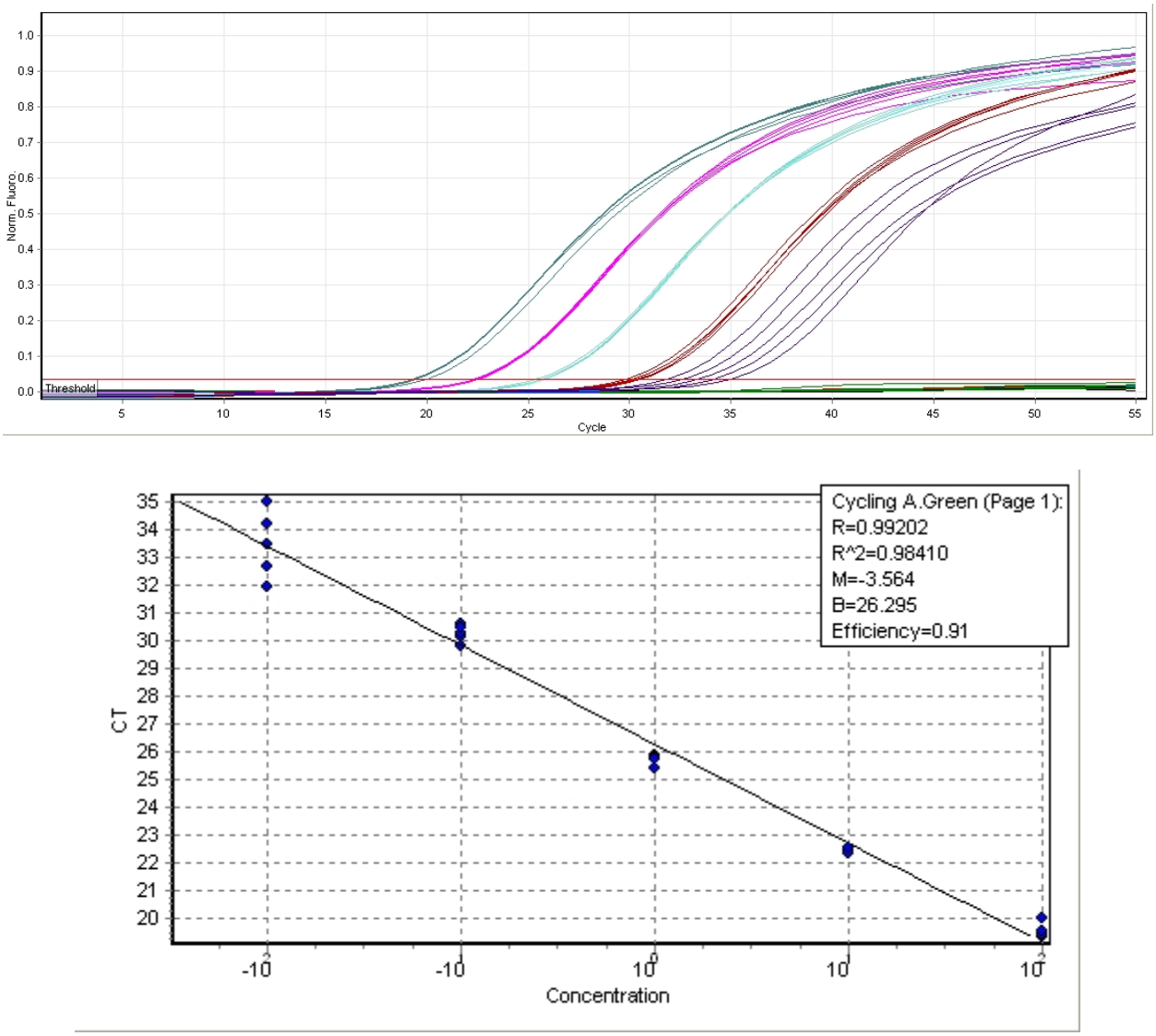

Figure 10. Amplification plot and corresponding standard curve of the *cp* gene of hermaphrodite-pollinated seeds at different total DNA concentrations. The amplification plot has five different colored curves (from left to right): 100 ng (green), 10 ng (pink), 1 ng (blue), 100 pg (brown), and 10 pg (purple) of total DNA. The parameters on the upper right are as follows, R: correlation coefficient, R^2: coefficient of determination, M: slope of the regression line, B: y-intercept of the regression line, Efficiency: calculated efficiency of the reaction.

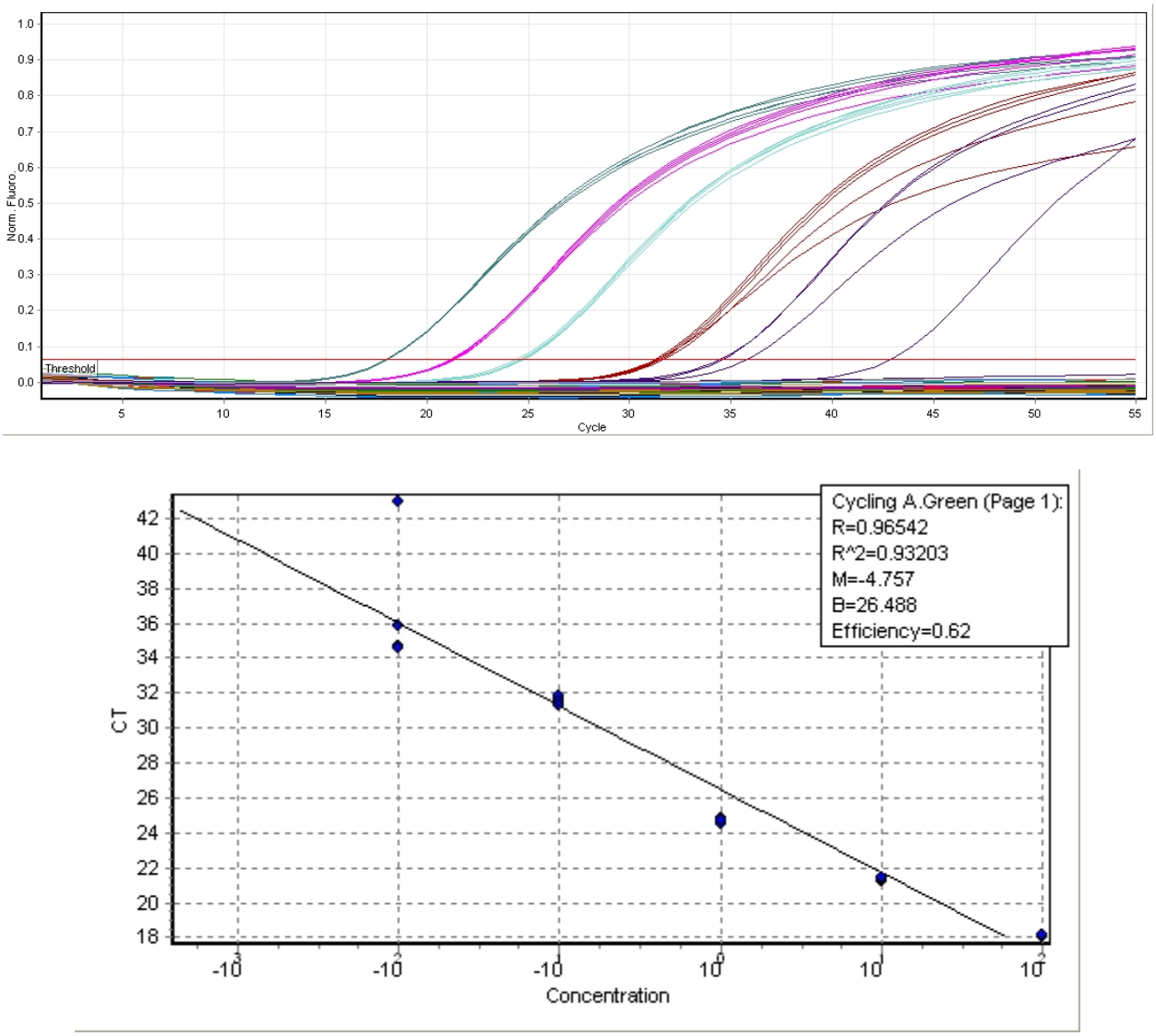

Figure 11. Amplification plot and corresponding standard curve of the *cp* gene of hermaphrodite leaves at different total DNA concentrations. The amplification plot has five different colored curves (from left to right): 100 ng (green), 10 ng (pink), 1 ng (blue), 100 pg (brown), and 10 pg (purple) of total DNA. The parameters on the upper right are as follows, R: correlation coefficient, R^2: coefficient of determination, M: slope of the regression line, B: y-intercept of the regression line, Efficiency: calculated efficiency of the reaction.

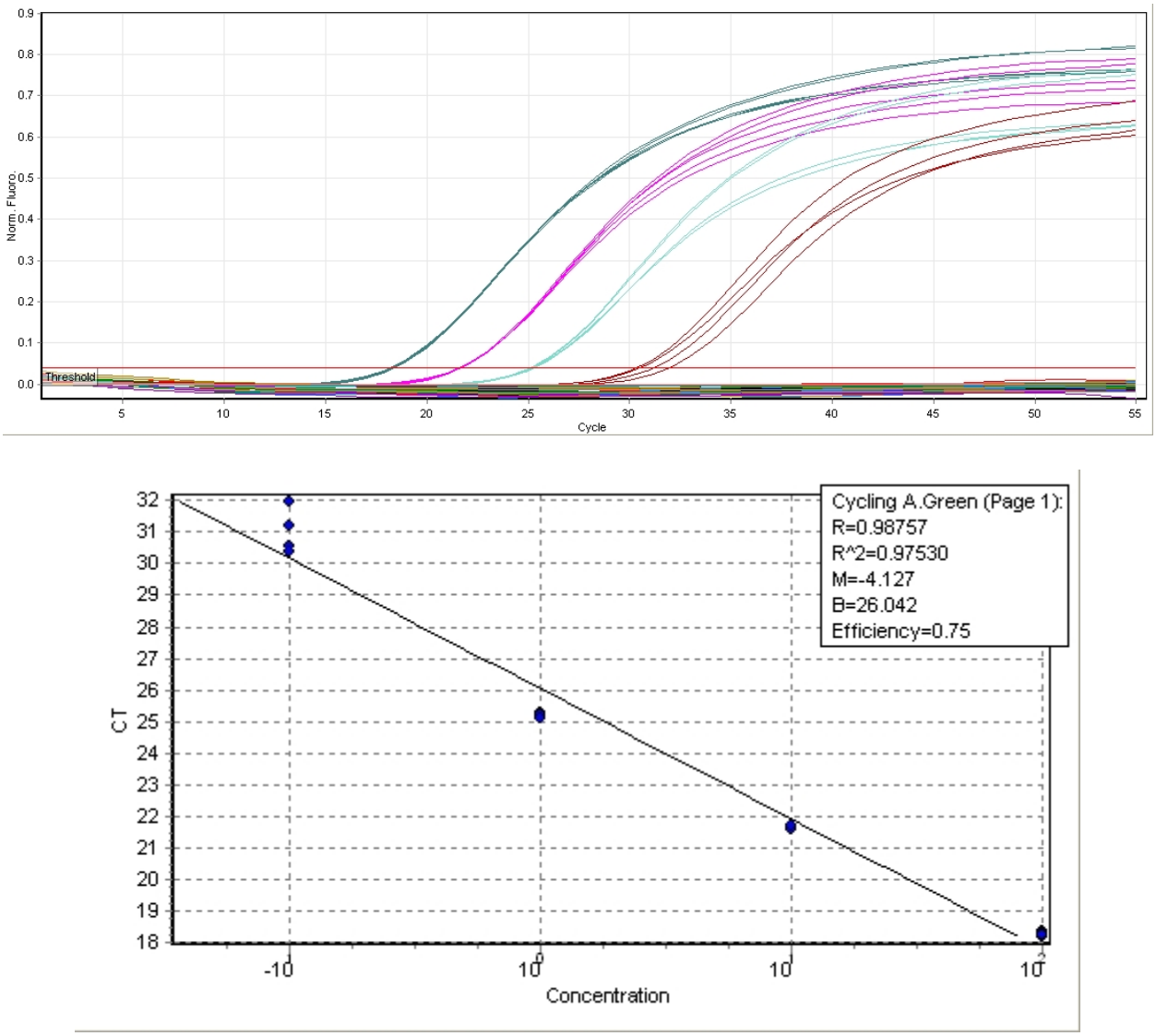

Figure 12. Amplification plot and corresponding standard curve of *male* marker of male leaves at different total DNA concentrations. The amplification plot has five different colored curves (from left to right): 100 ng (green), 10 ng (pink), 1 ng (blue), and 100 pg (brown) of total DNA. The parameters on the upper right are as follows, R: correlation coefficient, R^2: coefficient of determination, M: slope of the regression line, B: y-intercept of the regression line, Efficiency: calculated efficiency of the reaction.

Table 9. C_t value average of the five replicates of the *papain* gene and the target genes (*cp* and *male*). Values for the hermaphrodite-pollinated seed is in orange, hermaphrodite leaf is in blue, and male leaf is in green.

Concentration (ng/reaction)	***papain* Signal (C_t)**	***cp* signal (C_t)**	***male* signal (C_t)**
100	**19.29**	**21.24**	
10	**22.22**	**24.19**	
1	**25.24**	**27.46**	
0.1	**29.58**	**32.06**	
100	**17.19**	**18.11**	
10	**20.31**	**21.31**	
1	**23.84**	**24.72**	
0.1	**30.03**	**31.51**	
100	**17.27**		**17.67**
10	**20.34**		**21.05**
1	**23.75**		**24.59**
0.1	**28.93**		**30.40**

5.3.2 Calculating efficiencies for $2^{-\Delta\Delta Ct}$ assumption

To use the $2^{-\Delta\Delta Ct}$ method, the efficiencies of the three primer systems must be similar. Another way of saying this is that $C_{t,\,target} - C_{t,\,reference}$, or the ΔC_t, should remain constant within

samples over a range of sample DNA concentrations. To test this, ΔC_t values from the standards used to prepare the standard curves were regressed on their template concentrations, and the slope of the regression line was examined (Figure 13). A slope near zero means that ΔC_t remains nearly constant regardless of template concentration, because the primer efficiencies for target and reference genes are approximately equal. Thus, a slope near zero indicates that the data are appropriate for calculating target DNA percentages by the $2^{-\Delta\Delta Ct}$ method. The absolute values of the slopes of the lines in Figure 13 were different, but did not deviate much from the horizontal "zero" orientation. Livak and Schmittgen (2001) did not provide a numerical criterion to determine the boundaries of what was meant by "similar" efficiencies. However, these data, combined with the efficiency parameters (using the first three dilution levels) from the standard curves, suggest that the efficiencies of the three primer systems were similar enough to utilize the $2^{-\Delta\Delta Ct}$ method.

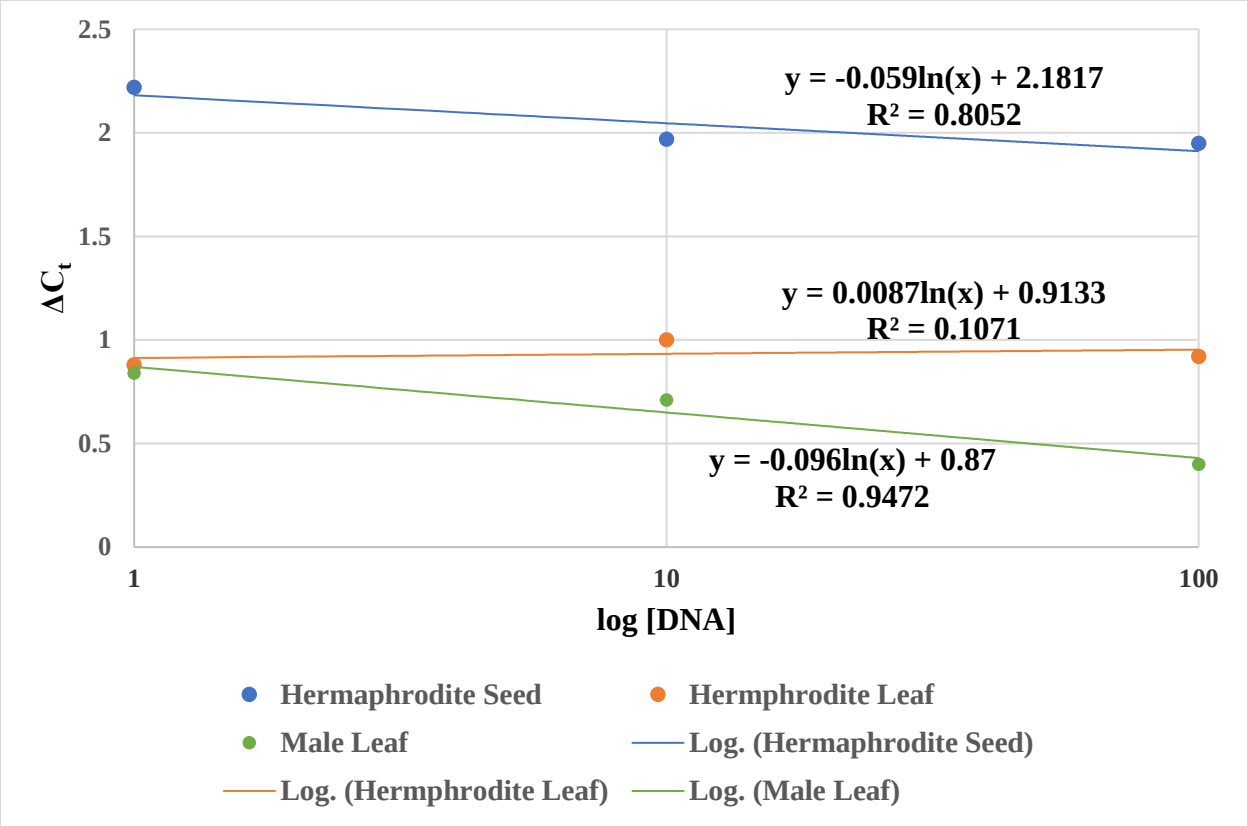

Figure 13. Linear regression of the log [DNA] and ΔC_t. This zero slope criterion is a prerequisite for use of the $2^{-\Delta\Delta Ct}$ method, because the calculation assumes the target and reference gene primers have similar reaction efficiencies.

5.3.3 Validation

The validation samples were created with known proportions of hermaphrodite and male leaf DNAs. The proportions tested were 1% hermaphrodite: 99% male, 10% hermaphrodite: 90% male, 90% hermaphrodite: 10% male, and 99% hermaphrodite: 1% male. The amplification curves for the mixtures showed an almost identical *papain* signal (Figure 14) which was an expected result, because the total papaya DNA was the same in each sample, regardless of the male to hermaphrodite ratio. The average of the *papain* signals from the four samples was calculated to be 17.55 and this value was used as the reference value for

normalization. The *coat protein* and *male* signals were quantified using the Rotor-Gene program and normalized using the $2^{-\Delta\Delta Ct}$ method (Table 10). This involved first normalizing samples for variation in the total genomic DNA concentration by calculating a ΔC_t value, the difference between the C_t average at each dilution of the target gene and 17.55, the average C_t of the reference gene (*papain*). The calculated ΔC_t for each target gene was then calibrated to the corresponding calibration point. This calibrator was chosen as the closest approximation to 100% target DNA existing among the target DNA mixtures used as validation templates. For the *cp* gene, the 99% hermaphrodite: 1% male sample was the calibrator. For the *male* marker, the 99% male: 1% hermaphrodite was the calibrator. The normalized and calibrated value, called the $\Delta\Delta C_t$, then, is the difference between the calibrator ΔC_t and the ΔC_t of the other samples for the same target gene. The $\Delta\Delta C_t$ value represents the number of amplification cycles needed to raise the concentration of target gene DNA in an unknown sample to the concentration existing concurrently in the 100% target calibration standard. Consequently, the $\Delta\Delta C_t$ of the 99% hermaphrodite mixture targeting the *cp* gene, and the 99% male mixture targeting the *male* marker, is 0. The $\Delta\Delta C_t$ values can be converted to fold difference in DNA concentration by calculating $2^{-\Delta\Delta Ct}$, since each cycle of amplification increased target DNA concentration two-fold. Finally, this fold difference can be related to the actual percentage of the target in the spiked DNA samples by the following formula:

$$100\ [(2^{-\Delta\Delta Ct})_{calibrator} / (2^{-\Delta\Delta Ct})_{sample}] = 100 / (2^{-\Delta\Delta Ct})_{sample} = \text{target gene \%}$$

The normalized and calibrated validation results were also compared to C_t values of the standard curve points. The standard points were adjusted similarly using the $2^{-\Delta\Delta Ct}$ method

(Table 11). The predicted target percentage for corresponding dilutions were similar for the adjusted validation and standard curve points. For example, comparing the *cp* gene percentage from 10% hermaphrodite: 90% male to the 10 ng/reaction (10% dilution of total genomic DNA) point on the standard curve, they both predicted approximately 12%. Thus, these results confirmed that the relative quantification method using the $2^{-\Delta\Delta Ct}$ calculation can reliably quantify the target percentages in the DNA template.

This assay can reliably detect and quantify target percentages as low as 1%. From a practical standpoint, with papaya fruits containing an average of 500 seeds, the assay can quantify targets consisting of as few as five seeds in the total seed bulk. However, the current assay is not as reliable below this target proportion. In the future, this assay should be improved upon by generating a standard curve with at least five reliable dilution points, compared to the three reliable points used in this thesis.

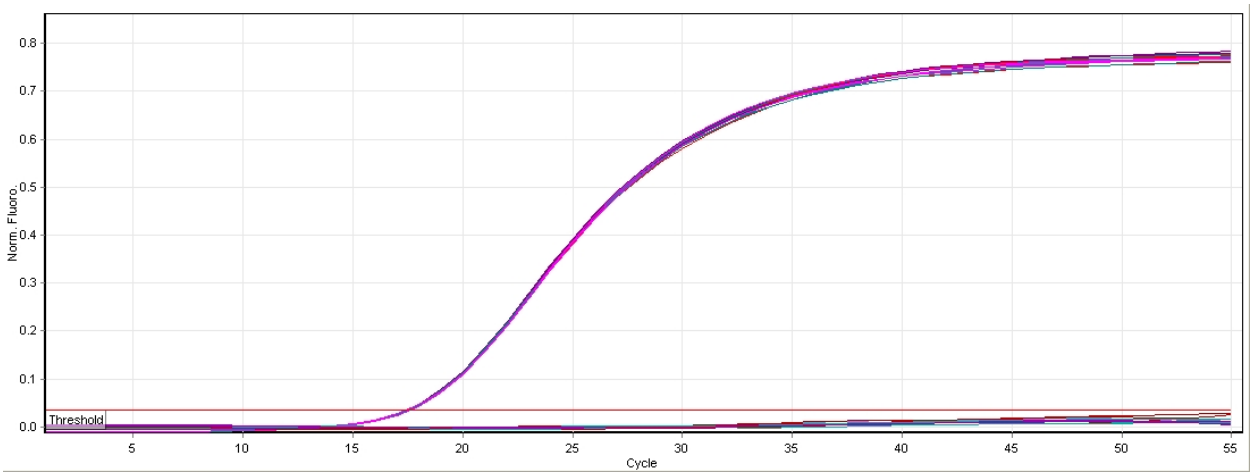

Figure 14. Amplification plot of the *papain* gene with four different proportions of male and hermaphrodite DNA.

Table 10. Normalized validation C_t values using the $2^{-\Delta\Delta Ct}$ method. The table is organized as follows: (a) *coat protein* results and (b) *male* results.

(a)

DNA Ratio	***papain*** **C_t**	***cp*** **C_t**	**ΔC_t**	**$\Delta\Delta C_t$**	**$2^{-\Delta\Delta Ct}$**	**Estimated *cp* composition (%)**
1% H: 99% M	**17.55**	**24.33**	**6.78**	**6.47**	**88.647**	**1.128**
10% H: 90% M		**20.88**	**3.33**	**3.02**	**8.112**	**12.327**
90% H: 10% M		**17.82**	**0.27**	**-0.04**	**0.973**	**102.775**
99% H: 1% M		**17.86**	**0.31**	**0**	**1**	**100**

(b)

DNA Ratio	***papain*** **C_t**	***male*** **C_t**	**ΔC_t**	**$\Delta\Delta C_t$**	**$2^{-\Delta\Delta Ct}$**	**Estimated male composition (%)**
1% H: 99% M	**17.55**	**18.1**	**0.55**	**0**	**1**	**100**
10% H: 90% M		**18.06**	**0.51**	**-0.04**	**0.973**	**102.774**
90% H: 10% M		**21.42**	**3.87**	**3.32**	**9.987**	**10.013**
99% H: 1% M		**24.86**	**7.31**	**6.76**	**108.383**	**0.92**

Table 11. Normalized standard curve C_t values using the $2^{-\Delta\Delta Ct}$ method. The table is organized as follows: (a) hermaphrodite-pollinated seed, (b) hermaphrodite leaf, and (c) male leaf.

(a)

[DNA] (ng)	***papain* C_t**	***cp* C_t**	**ΔC_t**	**$\Delta\Delta C_t$**	**$2^{-\Delta\Delta Ct}$**	**Estimated *cp* composition (%)**
100	**19.29**	**21.24**	**1.95**	**0**	**1**	**100**
10		**24.19**	**4.9**	**2.95**	**7.727**	**12.94**
1		**27.46**	**8.17**	**6.22**	**74.543**	**1.34**

(b)

[DNA] (ng)	***papain* C_t**	***cp* C_t**	**ΔC_t**	**$\Delta\Delta C_t$**	**$2^{-\Delta\Delta Ct}$**	**Estimated *cp* composition (%)**
100	**17.14**	**18.11**	**0.97**	**0**	**1**	**100**
10		**21.31**	**4.17**	**3.2**	**9.19**	**10.881**
1		**24.72**	**7.58**	**6.61**	**97.681**	**1.024**

(c)

[DNA] (ng)	***papain* C_t**	***male* C_t**	**ΔC_t**	**$\Delta\Delta C_t$**	**$2^{-\Delta\Delta Ct}$**	**Estimated male composition (%)**
100	**17.27**	**17.67**	**0.4**	**0**	**1**	**100**
10		**21.05**	**3.78**	**3.38**	**10.411**	**9.605**
1		**24.59**	**7.32**	**6.92**	**121.095**	**0.826**

5.4 Conclusion and future work

In conclusion, this thesis developed a probe-based qPCR methodology that allows for the quantitation of the portion of bulked seed samples that are attributable to pollination by male and hermaphrodite pollen sources. Primer and probe sequences of the *papain* gene, *coat protein* gene, and *male* marker were developed and described in this thesis. The primers and probe used to target male-specific sequence was successful, evidenced by gel electrophoresis and Sanger sequencing. Several extraction methodologies were tested, improved, and optimized for both papaya seed and leaf tissue. The methodology developed during this thesis project can extract DNA from up to 500 dry papaya seeds. The DNA purity was adequate for efficient amplification during PCR.

The qPCR analyses of the three genes (or markers) were successful, but there are many areas for improvement. Briefly, this thesis was successful in developing reproducible standard curves for the three markers, which could estimate the percentage of target gene relative to the total genomic DNA concentration. Validation using known percentages of male and hermaphrodite DNA were tested and the amplification results could estimate the target gene percentage with fair accuracy. Analysis using the $2^{-\Delta\Delta Ct}$ method were based on three levels of DNA dilutions (100 ng to 1 ng). To improve on the reliability and accuracy of the assay, it is recommended to further optimize the protocol to obtain five points on the standard curve, so that the regression is reliable over a larger range of target concentrations.

The methodology developed in this thesis will be used to assay the seeds from the experimental plot discussed in Chapter 1 (Figure 2). By extracting DNA from all the seeds in a single fruit, the relative pollen contribution from male and hermaphrodite can be determined. Hopefully, these data and the methodology will aid in future studies of papaya reproductive

biology and in formulating biocontainment methods for experimentation with regulated transgenic papaya.

Appendix A. Genomic Sequences from BLAST Search

Carica papaya papain mRNA, complete cds (GenBank No.: M15203.1)

The sequence below is the annealing sites of the *papain* primer. The forward primer, probe, and reverse primer sequences are noted below.

```
   1 atccattccc acttaagaag taaaaagata tagctagtgt cacaatggct atgatacctt
  61 caatttcaaa gttgcttttt gttgcaatat gtctttttgt ttatatgggt ttgtcatttg
 121 gtgatttttc tattgtgggt tattctcaaa atgacttgac atccactgaa agacttattc
 181 agctatttga atcgtggatg ttgaagcaca ataagattta taagaatatt gatgaaaaaa
 241 tctacagatt tgaaattttt aaagataatc ttaaatatat tgatgagaca aataagaaaa
 301 ataacagtta ttggcttgga ttaaatgtgt ttgctgatat gagcaatgat gaattcaaag
 361 aaaagtatac tggttctatt gctggaaatt atacaacaac cgaactatca tacgaagaag
 421 tgcttaatga tggtgatgta aatatcccgg agtatgtcga ttggagacaa aaaggagctg
 481 tcactcctgt aaaaaatcag ggttcttgtg gtagttgttg ggcattctca gctgttgtaa
 541 ctatagaggg aataattaag attagaactg ggaacttaaa tgaatactca gagcaagaac
 601 tgcttgactg cgacagacgt agctacgggt gcaatggagg ttacccttgg agtgcacttc
 661 aattagtggc tcaatatggt attcactaca gaaatactta cccatatgag ggagtgcaac
 721 gttattgtcg ctcaagggag aaaggtcctt atgcagccaa aaccgatggg gttcgacaag
 781 tgcaaccata taatgaaggg gctctcttat attcaattgc aaaccaacct gtgagcgttg
 841 tccttgaagc tgctggaaaa gatttccaat tatatagagg gggaatattt gtggggccat
 901 gcggaaacaa agtagatcat gcagtcgcag cagttgggta tggaccaaat tacatactca
 961 taaagaattc atggggtaca ggatggggtg aaaatggata tataagaatc aaaagaggca
1021 ctggaaactc ctatggagta tgtggacttt atacaagctc attctatcct gttaaaaact
1081 gatgagatca cggctttcat aaaatccctt atatatatat atatatatag aacttgtata
1141 ctcatcgtgt gttgaaataa taaatgagag gattaataat ttgtgtaatc ctatatatca
1201 cagtttgttg tgacaaactt ttgcatcgct tgttatatta tttgtaataa tgtttgtttt
1261 gattgaataa acttttacat atacttttat gc
```

Carica papaya transgenic cultivar Rainbow beta-glucuronidase, coat protein… (GenBank No: FJ467933.1)

The sequence is a portion of the transgene that was transformed into 'Rainbow' papaya. The forward primer, probe, and reverse primer sequences obtained from Xu et al. (2008) is shown below. There are three mismatches in the probe and two mismatches in the reverse primer sequences, noted in red font.

```
9121 ttcttccttg ttactaacac tgccgtccat accaaacatt ttgcgactgg tgtttcgcag
9181 cgctgcagcc ttcatctgca tgtgagcttc gcgagcccta tcaggtgttt tcgaattcac
9241 ctcatagaag tcgaaagcgt atctagcgag gctaatgtca gtcaaatttc tcttgattcc
9301 ataccgcggc atgtacctct cagtagcatt tctcttcgca atgtatgctt ctgccgcgtt
9361 actaaagtga gccataattt gcctaaatga cggagtagca tgctcaatca aaggcttgat
9421 tggataatca acttgggttt ccccatccat cataacccag acaccagata tgtctggaga
```

Appendix B. List of primer candidates from *Primer 3 Plus* software

The following is the list of 18 primer candidates that were mentioned in Chapter 3.

Pair 1 (Low Temp)

Forward AGCCCATGCACATGTGTGGTCAA

Probe ACACACCCTGCGTGCATACGAAAATGCGCAGGA

Reverse TCGTGTTGTTAGGCCAAGTGGGT

Pair 2 (Low Temp)

Forward AGCCCATGCACATGTGTGGTCA

Probe ACACACCCTGCGTGCATACGAAAATGCGCAGGA

Reverse TCGTGTTGTTAGGCCAAGTGGGT

Pair 3 (Low Temp)

Forward TCAGCCCATGCACATGTGTGGT

Probe ACACACCCTGCGTGCATACGAAAATGCGCAGGA

Reverse TCGTGTTGTTAGGCCAAGTGGGT

Pair 4 (Low Temp)

Forward CATGCACATGTGTGGTCAAGGCA

Probe ACACACCCTGCGTGCATACGAAAATGCGCAGGA

Reverse TCGTGTTGTTAGGCCAAGTGGGT

Pair 5 (Low Temp)

Forward GCCCATGCACATGTGTGGTCAA

Probe ACACACCCTGCGTGCATACGAAAATGCGCAGGA

Reverse TCGTGTTGTTAGGCCAAGTGGGT

Pair 6 (Low Temp)

Forward CAGCCCATGCACATGTGTGGTCA

Probe ACACACCCTGCGTGCATACGAAAATGCGCAGGA

Reverse TCGTGTTGTTAGGCCAAGTGGGT

Pair 7 (Low Temp)

Forward CCATGCACATGTGTGGTCAAGGC

Probe ACACACCCTGCGTGCATACGAAAATGCGCAGGA

Reverse TCGTGTTGTTAGGCCAAGTGGGT

Pair 8 (Low Temp)

Forward GCCCATGCACATGTGTGGTCAAG

Probe ACACACCCTGCGTGCATACGAAAATGCGCAGGA

Reverse TCGTGTTGTTAGGCCAAGTGGGT

Pair 1 (High Temp)

Forward AAACGAGTAACCAAGCCCACAC

Probe GGTCCATGGCATGCCTACAAGCATTCGAGAAAGCA

Reverse AATGCCTTGACCACACATGTGC

Pair 2 (High Temp)

Forward GCACTTACACATTCAGCCCATGC

Probe TGTGTGGTCAAGGCATTCTACACACCCTGCGT

Reverse TCTCCTGCGCATTTTCGTATGC

Pair 3 (High Temp)

Forward ACATTCAGCCCATGCACATGTG

Probe TGGTCAAGGCATTCTACACACCCTGCGTGCA

Reverse TGTCGCATATATGTGGGGCTTC

Pair 4 (High Temp)

Forward TTCCCCACCTGCATAGCCAAG

Probe AACCAAGCCCACACGGTCCATGGCATGCCT

Reverse ACATGTGCATGGGCTGAATGTG

Pair 5 (High Temp)

Forward CGTGTGAACTATGGCCTATCACG

Probe TGGGTTCCCCACCTGCATAGCCAAGCATACA

Reverse TGTGGGCTTGGTTACTCGTTTG

Pair 6 (High Temp)

Forward TCCATGGCATGCCTACAAGC

Probe AGCACTTACACATTCAGCCCATGCACATGTGTGGT

Reverse ACGCAGGGTGTGTAGAATGC

Pair 7 (High Temp)

Forward GCCTACAAGCATTCGAGAAAGC

Probe TGGTCAAGGCATTCTACACACCCTGCGTGCA

Reverse TCGCATATATGTGGGGCTTCTCC

Pair 8 (High Temp)

Forward ATGGCCTATCACGTATGGGTTCC

Probe ACGAGTAACCAAGCCCACACGGTCCATGGC

Reverse GCTTTCTCGAATGCTTGTAGGC

Pair 9 (High Temp)

Forward ACCTGCATAGCCAAGCATACAAC

Probe AACCAAGCCCACACGGTCCATGGCATGCCT

Reverse GCATGGGCTGAATGTGTAAGTGC

Pair 10 (High Temp)

Forward ACGGTCCATGGCATGCCTAC

Probe TGGTCAAGGCATTCTACACACCCTGCGTGCA

Reverse GGCTTCTCCTGCGCATTTTCG

Appendix C. Reagents

Chloroform: isoamyl alcohol (24:1)

Combine 800 µL isoamyl alcohol and 24.2 mL chloroform (Thermo Fischer Scientific, Inc., Waltham, MA, USA) to make a 25-mL stock of chloroform: isoamyl alcohol (24:1).

CTAB Buffer for "Rao" Protocol (10 mL buffer: 1 gram of seeds)

The volume of CTAB buffer required is dependent on the mass of seeds used for the extraction. The buffer is made with the following: 100 mM Tris-HCl, 20 mM EDTA (pH 8.0, Sigma Aldrich), 1.4 M NaCl, 2% CTAB (Sigma Aldrich), 1% PVP-40 (Sigma Aldrich), 1% PVPP-40 (Sigma Aldrich), and 2% 2-mercaptoethanol (Sigma Aldrich). It is highly recommended to add the 2-mercaptoethanol moments before adding the seed tissue into the buffer.

CTAB buffer for "Matsumoto" Protocol (100 mL)

The components of this buffer are: 3% CTAB (add 3 grams), 1.4 M NaCl (add 8.2 grams), 100 mM Tris-HCl, 20 mM EDTA (pH 8.0), 0.2% 2-mercaptoethanol (add 200 µL), 1% PVP-40 (add 1.0 grams). For the TE buffer, either use 100 mL at the concentration in the previous sentence, or use 10 mL of 1 M Tris-HCl, 200 mM EDTA (pH 8.0) with 90 mL of deionized water. Add the 2-mercaptoethanol to the buffer just prior to mixing the buffer with the seeds.

70% ethanol

Combine 737 mL of 95% ethanol with 263 mL of deionized water to make a 1-L stock of 70% ethanol.

Isopropanol

Store isopropanol used for CTAB extraction in a -20°C freezer.

Phenol: chloroform: isoamyl alcohol (25:24:1)

Purchased from Sigma Aldrich (St. Louis, MO, USA).

RNase A (10 mg/mL)

Came with the DNeasy Plant Mini Kit (Qiagen Inc., Valencia, CA, USA).

Rotor-Gene SYBR Green PCR Kit

Purchased from Qiagen, Inc. (Valencia, CA, USA). The master mix contains HotStarTaq *Plus* DNA Polymerase, Rotor-Gene SYBR Green PCR buffer, and dNTP mix.

Rotor-Gene Probe PCR Kit

Purchased from Qiagen, Inc. (Valencia, CA, USA). The master mix contains HotStarTaq *Plus* DNA Polymerase, Rotor-Gene SYBR Green PCR buffer, and dNTP mix.

3.0 M sodium acetate

Add 12.3 grams of sodium acetate in 50 mL of ultra-pure water.

7.5 M sodium acetate

Add 615 mg of sodium acetate in 1 mL of ultra-pure water.

5x TBE Buffer (pH 8.0)

Add 54 grams of Tris-HCl (Sigma Aldrich), 27.5 grams of boric acid, 3.7 grams of EDTA (Sigma Aldrich), and 0.5 grams of NaOH to 1 L of deionized water. Adjust the pH to 8.0.

1 M Tris-HCl, 200 mM EDTA (pH 8.0)

Add 88.8 grams of Tris-HCl, 53.2 grams Trizma-base (Sigma Aldrich), 67.2 grams of EDTA to approximately 900 mL of deionized water. Adjust the pH to 8.0, then add deionized water to reach 1 L of solution.

100 mM Tris-HCl, 10 mM EDTA (pH 8.0)

Add 266.4 mg of Tris-HCl, 159 mg Trizma-base, and 100.86 mg EDTA to approximately 20 mL of deionized water. Adjust the pH to 8.0, then add deionized water to reach 30 mL of solution.

100 mM Tris-HCl, 20 mM EDTA (pH 8.0)

Add 266.4 mg of Tris-HCl, 159 mg Trizma-base, and 200.86 mg EDTA to approximately 20 mL of deionized water. Adjust the pH to 8.0, then add deionized water to reach 30 mL of solution.

Appendix D. Papaya Seed DNA Extraction "Rao" Protocol

Overview

This protocol is adapted from a paper by Nageswara-Rao et al. (2013). This procedure can also be used for papaya leaf tissue (attempted for this thesis), although it is not mentioned in the literature. For this procedure, it is optimal to use 300 to 500 papaya seeds, which is approximately 5.0 to 9.0 grams. For papaya leaf, it is recommended to use approximately 5.0 grams of tissue. For this thesis, 500 papaya seeds were used.

Materials

-500 Papaya Seeds

-Mortar & Pestle

-Micropipettes (1,000, 200, 100, 10, and 2 µL)

-Water bath (set at 65°C)

-Parafilm

-Centrifuge compatible with 50-mL Falcon tubes (at least 1,000*g*)

-Centrifuge compatible with 2-mL tubes (6,000 to 20,000*g* recommended)

-Centrifuge compatible with 2-mL tubes that can be temperature controlled to 4°C

-One 250-mL beaker

-One 50-mL conical Falcon tube

-Five 2-mL centrifuge tubes (label them A through E)

Reagents

1. **Stock #1**: 250 mL of 1 M Tris-HCl, 200 mM EDTA (pH 8.0)
2. **Stock #2**: 30 mL of 100 mM Tris-HCl, 10 mM EDTA (pH 8.0)
3. Approximately 50 mL of CTAB Buffer (10 mL buffer: 1 gram of seeds)
4. 2 mL of phenol: chloroform: isoamyl alcohol (25:24:1)
5. 2 mL of chloroform: isoamyl alcohol (24:1)
6. 2.4 mL of chilled isopropanol
7. 70 uL of 70% ethanol
8. 150 uL of 10 mM Tris-HCl, 1 mM EDTA (pH 8.0)
9. 1 uL of RNase A (10 mg/mL)
10. 60 uL of 7.5 M sodium acetate

Procedure

1. Turn the water bath on and set it to 65°C.

2. Weigh the seed batch and calculate the appropriate volume of CTAB buffer to make. Seed yields from a single papaya fruit can vary, but it is often 300 to 500 seeds. It is recommended to make about 50 mL of the buffer in a 250-mL beaker. Apply Parafilm over the beaker and incubate the buffer at 65°C.

3. Grind the seeds with a mortar and pestle under liquid nitrogen. Do not allow the tissue to thaw.

4. Add 2-mercaptoethanol to the CTAB buffer. Then, scrape the frozen seed tissue into the buffer. Mix the mixture thoroughly using a magnetic stirrer. Cover the beaker once again

and incubate at 65°C for 45 minutes, shaking the solution every 5 to 10 minutes. After this step, reduce the water bath temperature to 37°C for step 11.

5. Pour the solution into a 50-mL falcon tube. If the volume is higher, use two tubes. Spin the tube at room temperature for 15 minutes at 2,000*g*.

6. Transfer 1 mL of the supernatant to a 2-mL centrifuge tube "A". Add 1 mL of phenol: chloroform: isoamyl alcohol (25:24:1). Mix gently by inversion and centrifuge at 10,000*g* for 10 minutes at room temperature. After the spin, the solution will separate into three layers: aqueous layer on the top, debris and proteins in the middle, and the organic layer on the bottom. Transfer the entire aqueous layer to tube "B" and add an equal volume of phenol: chloroform: isoamyl alcohol (25:24:1). Mix and centrifuge at 10,000*g* for 10 minutes at room temperature.

7. Transfer the entire aqueous layer to tube "C" and an equal volume of chloroform: isoamyl alcohol (24:1). Mix and centrifuge at 10,000*g* for 10 minutes at room temperature. Transfer the aqueous layer to tube "D" and add and equal volume of chloroform: isoamyl alcohol (24:1). Mix and centrifuge at 10,000*g* for 10 minutes at room temperature.

8. Transfer 600 µL of the aqueous layer to tube "E" and precipitate the DNA with two volumes (1,200 µL) of chilled isopropanol. Set solution in a -20°C freezer for 30 minutes. Centrifuge the tube at 10,000*g* for 10 minutes at 4°C.

9. Discard the supernatant and wash the DNA pellet with 50 µL of 70% ethanol. Centrifuge the tube at 10,000g for 5 minutes at 4°C. Discard the supernatant and dry the pellet for 10 minutes.

10. Suspend the pellet in 100 µL of TE [10 mM Tris-HCl, 1 mM EDTA (pH 8.0)] buffer. Mix gently and allow the pellet to solubilize before continuing to the next step.

11. Add 1 µL of RNase A (10 mg/mL), cap the tube and vortex gently. Incubate the tube at 37°C for 30 minutes.

12. Re-precipitate the DNA by adding 1,200 µL of chilled isopropanol and 60 µL of 7.5 M sodium acetate. Incubate at -20°C for 30 minutes to 1 hour. Centrifuge at 10,000*g* for 15 minutes set at 4°C.

13. Discard the supernatant and add 20 µL of 70% ethanol. Mix the solution gently and centrifuge at 10,000g for 10 minutes set at 4°C.

14. Discard the supernatant and dry the pellet. Re-suspend the pellet in 50 µL of TE buffer.

Appendix E. Papaya Seed DNA Extraction "Matsumoto" Protocol

Overview

This protocol is an improved and optimized procedure mentioned in a paper by Matsumoto et al. (2010). It combines a popular CTAB based extraction method and the DNeasy Plant Mini Kit produced by QIAGEN, Inc. This procedure is specific for papaya seeds and its viability with papaya leaf tissue has not been tested. The procedure uses 500 papaya seeds as the starting unit and uses 100 mL of CTAB buffer regardless of the mass.

This procedure was discovered after the "Rao" protocol and since has replaced the "Rao" protocol. The advantages of the "Matsumoto" protocol, in comparison to the "Rao" protocol, is that it is much faster and the yield does not differ much from the "Rao" protocol despite a higher suspension volume. Also, this entire procedure can be performed at room temperature.

Materials

-500 Papaya Seeds

-Mortar & Pestle

-Micropipettes (1,000, 200, 100, 10, and 2 µL)

-Water bath (set at 65°C)

-Parafilm

-Miracloth

-Centrifuge compatible with 50-mL Falcon tubes (at least 1,000*g*)

-Centrifuge compatible with 2-mL tubes (6,000 to 20,000*g* recommended)

-One 250-mL beaker & Three 50-mL conical Falcon tubes

Reagents

1. CTAB Buffer (100 mL)
2. Chloroform: isoamyl alcohol (24:1; 20 mL)
3. Chilled isopropanol (approximately 20 mL)
4. 70% ethanol (2 mL)
5. TE buffer 1: [1 M Tris-HCl, 200 mM EDTA (pH 8.0); 10 mL]
6. TE buffer 2: [100 mM Tris-HCl, 20 mM EDTA (pH 8.0); 400 µL]
7. All reagents included in DNeasy Plant Mini Kit (QIAGEN)

Procedure

CTAB Section

1. Turn the water bath on and set it to 65°C.
2. Make the CTAB buffer in a 250-mL beaker. Apply Parafilm over the beaker and incubate the buffer at 65°C.
3. Grind the seeds with a mortar and pestle under liquid nitrogen.
4. Add 2-mercaptoethanol to the CTAB buffer. Scrape the frozen seed tissue into the buffer. Mix thoroughly using a magnetic stirrer. Cover the beaker once again with Parafilm and incubate at 65°C for 30 minutes, shaking the mixture every 5 minutes.
5. Place a Miracloth in a funnel and attach it to an open 50-mL Falcon tube "A". Gently pour the CTAB-seed mixture on top of the Miracloth and collect 45 mL of solution in the tube.
6. Centrifuge the tube at 2,000*g* for 15 minutes at room temperature.

7. Transfer 30 mL of the supernatant to a new tube “B”. Add 20 mL of chloroform: isoamyl alcohol (24:1). Close the cap, gently invert several times, and centrifuge at 2,000*g* for 15 minutes at room temperature.
8. Transfer 25 mL of the top aqueous layer to a new tube “C”. Add 2/3 volume (approximately 17 mL) of chilled isopropanol and 2.5 mL of 3 M sodium acetate. Gently mix and incubate the tube on ice for 30 minutes.
9. Spin the tube at 2,000*g* for 15 minutes at room temperature.
10. Decant the supernatant and transfer the pellet to a 2-mL tube. Add 2 mL of 70% ethanol to the pellet. Close tube, gently disturb the pellet, and centrifuge at 10,000*g* for 15 minutes.
11. Carefully pipette out the supernatant and dry the pellet for 15 minutes. Then, resuspend the pellet in 400 µL of TE buffer 2. Often, the pellet will not dissolve readily so light shaking and vortex is recommended. It is also advised to allow the resuspension in the refrigerator (4°C).
12. Transfer 100 µL of the DNA solution to a new 2-mL tube. This solution becomes the first step of the QIAGEN DNeasy Plant Mini Kit. It is recommended to use the remaining 300 µL of the DNA solution and do the same procedure. This will increase the yield of DNA at the end.

<u>QIAGEN DNeasy Kit Section (please see the manual for more details)</u>

13. Add 400 µL of AP1 and 4 µL RNase A (100 mg/mL) to the 100 µL of DNA from the previous step. Incubate the tube at 65°C for 10 minutes, with agitation every 2 to 3 minutes.

14. Add 130 µL of buffer P3, mix the solution, and incubate the tube on ice for 5 minutes.
15. Centrifuge the mixture at 20,000*g* for 5 minutes at room temperature.
16. Add the supernatant to the lilac-colored QIAshredder column. Centrifuge at 20,000*g* for 2 minutes at room temperature.
17. Transfer the flow-through to a new 2-mL tube and keep track of the volume transferred in this step. Usually, about 600 µL is transferred at this step and it is recommended to use a 1,000 µL pipette set at 500 µL and a 20 µL pipette set at 20 µL to get the exact volume.
18. Add 1.5 volumes (of the previous transfer) of AW1 buffer and mix immediately by pipetting. Take 650 µL from this mixture and transfer it to the white-colored DNeasy spin column. Centrifuge at 6,000*g* for 1 minute.
19. Discard the supernatant from the collection tube. Repeat the 650 µL transfer until all the solution from step #18 has gone through the column. Discard the collection tube at the end (and the flow-through) and put the DNeasy column in a new collection tube.
20. Add 500 µL of buffer AW2 on to the column. Centrifuge at 6,000*g* for 1 minute. Discard the flow-through, but reuse the collection tube. Add 500 µL of buffer AW2 once again and centrifuge similarly. Discard both the flow-through and the collection tube.
21. Transfer the spin column to a 2-mL tube. Add 75 µL of buffer AE onto the spin column. Incubate at room temperature for 5 minutes. Centrifuge the tube for 1

minute at 6,000*g*. Add 60 uL of buffer AE onto the spin column. Incubate at room temperature for 5 minutes. Centrifuge the tube for 1 minute at 6,000*g*.

REFERENCES

ABM. (n.d.). PCR - Introduction | ABM Inc. Retrieved from https://www.abmgood.com/marketing/knowledge_base/polymerase_chain_reaction_introduction.php

Ahmed, F. E. (2002). Detection of genetically modified organisms in foods. *Trends in Biotechnology, 20*(5), 215-223. doi:http://dx.doi.org/10.1016/S0167-7799(01)01920-5

Antunes Carvalho, F., & Renner, S. S. (2012). A dated phylogeny of the papaya family (Caricaceae) reveals the crop's closest relatives and the family's biogeographic history. *Molecular Phylogenetics and Evolution, 65*(1), 46-53. doi:10.1016/j.ympev.2012.05.019

Bio-Synthesis, I. (n.d.). TaqMan® vs. SYBR® Green Chemistries.

Brock, T. D., & Freeze, H. (1969). Thermus aquaticus gen. n. and sp. n., a Nonsporulating Extreme Thermophile. *The Journal of Bacteriology, 98*(1), 289.

Bubner, B., & Baldwin, I. (2004). Use of real-time PCR for determining copy number and zygosity in transgenic plants. *Plant Cell Reports, 23*(5), 263-271. doi:10.1007/s00299-004-0859-y

Candolle, A. d. (1885). *Origin of cultivated plants*. New York: New York, D. Appleton and company.

Chandrika, U. G., Jansz, E. R., Wickramasinghe, S. M. D. N., & Warnasuriya, N. D. (2003). Carotenoids in yellow‐ and red‐fleshed papaya (Carica papaya L). *Journal of the Science of Food and Agriculture, 83*(12), 1279-1282. doi:10.1002/jsfa.1533

Chen, C., Yu, Q., Hou, S., Li, Y., Eustice, M., Skelton, R. L., . . . Ming, R. (2007). Construction of a sequence-tagged high-density genetic map of papaya for comparative structural and evolutionary genomics in Brassicales.(Report). *Genetics, 177*(4), 2481. doi:10.1534/genetics.107.081463

Clarke, J. D. (2009). Cetyltrimethyl ammonium bromide (CTAB) DNA miniprep for plant DNA isolation. *Cold Spring Harbor protocols, 2009*(3), pdb.prot5177. doi:10.1101/pdb.prot5177

Decraene, L. P. R., & Smets, E. F. (1999). The floral development and anatomy of Carica papaya (Caricaceae). *Canadian Journal of Botany, 77*(4), 582-598. doi:10.1139/b99-026

Deputy, J. C., Ming, R., Ma, H., Liu, Z., Fitch, M. M. M., Wang, M., . . . Stiles, J. I. (2002). Molecular markers for sex determination in papaya (Carica papaya L.). *International Journal of Plant Breeding Research, 106*(1), 107-111. doi:10.1007/s00122-002-0995-0

Doyle, J. J., & Doyle, J. L. (1987). A rapid isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin, 19*, 11-15.

Fang, G., Hammar, S., & Grumet, R. (1992). A quick and inexpensive method for removing polysaccharides from plant genomic DNA. *BioTechniques, 13*(1), 52.

FAOSTAT. (2014). *World Crop Production*. Retrieved from: http://faostat.fao.org/site/567/DesktopDefault.aspx?PageID=567#anco

Fraley, R. T., Rogers, S. G., Horsch, R. B., Sanders, P. R., Flick, J. S., Adams, S. P., . . . Woo, S. C. (1983). Expression of Bacterial Genes in Plant Cells. *Proceedings of the National Academy of Sciences of the United States of America, 80*(15), 4803-4807.

Ginzinger, D. G. (2002). Gene quantification using real-time quantitative PCR: An emerging technology hits the mainstream. *Experimental Hematology, 30*(6), 503-512. doi:10.1016/S0301-472X(02)00806-8

Gonsalves, D. (1998). CONTROL OF PAPAYA RINGSPOT VIRUS IN PAPAYA: A Case Study. *Annual Review of Phytopathology, 36*(1), 415-437. doi:10.1146/annurev.phyto.36.1.415

Guo, J., Yang, L., Liu, X., Zhang, H., Qian, B., & Zhang, D. (2009). Applicability of the chymopapain gene used as endogenous reference gene for transgenic huanong no. 1 papaya detection. *Journal of Agricultural and Food Chemistry, 57*(15), 6502. doi:10.1021/jf900656t

Hanania, U., Velcheva, M., Sahar, N., & Perl, A. (2004). An improved method for isolating high-quality DNA from Vitis vinifera nuclei. *Plant Molecular Biology Reporter, 22*(2), 173-177. doi:10.1007/BF02772724

Heid, C. A., Stevens, J., Livak, K. J., & Williams, P. M. (1996). Real time quantitative PCR. *Genome research, 6*(10), 986.

Jensen, D. D. (1949). *Papaya virus diseases with special reference to papaya ringspot*. Worcester, Mass.: Worcester, Mass.

Liao, Z., Yu, Q., & Ming, R. (2017). Development of male-specific markers and identification of sex reversal mutants in papaya. *International Journal of Plant Breeding, 213*(2), 1-12. doi:10.1007/s10681-016-1806-z

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. *Methods, 25*(4), 402-408. doi:10.1006/meth.2001.1262

Lyamichev, V., & Dahlberg, J. E. (1993). Structure-Specific Endonucleolytic Cleavage of Nucleic Acids by Eubacterial DNA Polymerases. *Science, 260*(5109), 778-783.

Ma, H., Moore, P. H., Liu, Z., Kim, M. S., Yu, Q., Fitch, M. M. M., . . . Ming, R. (2004). High-density linkage mapping revealed suppression of recombination at the sex determination locus in papaya. *Genetics, 166*(1), 419.

Manshardt, R. (1999). *UH Rainbow papaya : a high-quality hybrid with genetically engineered disease resistance* (Rev.. ed.). Honolulu: Honolulu : Cooperative Extension Service, College of Tropical Agriculture and Human Resources, University of Hawaii at Manoa.

Matsumoto, T. K., Zee, F. T. P., Suzuki, J. Y., Tripathi, S., & Carr, J. (2010). Determining Sex and Screening for the Adventitious Presence of Transgenic Material in *Carica papaya* L. Seed Germplasm. *Horticultural Science, 45*(1), 161-164.

Ming, R., Hou, S., Feng, Y., Yu, Q., Dionne-Laporte, A., Saw, J. H., . . . Alam, M. (2008). The draft genome of the transgenic tropical fruit tree papaya (Carica papaya Linnaeus). *Nature, 452*(7190), 991. doi:10.1038/nature06856

Ming, R., Yu, Q., & Moore, P. H. (2007). Sex determination in papaya. *Seminars in Cell and Developmental Biology, 18*(3), 401-408. doi:10.1016/j.semcdb.2006.11.013

Morrison, T. B., Weis, J. J., & Wittwer, C. T. (1998). Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification. *BioTechniques, 24*(6), 954.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., & Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbor symposia on quantitative biology, 51*, 263-273.

Murray, M. G., & Thompson, W. F. (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic acids research, 8*(19), 4321.

Na, J.-K., Wang, J., Murray, J. E., Gschwend, A. R., Zhang, W., Yu, Q., . . . Ming, R. (2012). Construction of physical maps for the sex-specific regions of papaya sex chromosomes. *BMC genomics, 13*, 176. doi:10.1186/1471-2164-13-176

Nageswara-Rao, M., Kwit, C., Agarwal, S., Patton, M. T., Skeen, J. A., Yuan, J. S., . . . Stewart, C. N. (2013). Sensitivity of a real-time PCR method for the detection of transgenes in a mixture of transgenic and non-transgenic seeds of papaya (Carica papaya L.). *BMC biotechnology, 13*, 69. doi:10.1186/1472-6750-13-69

NASS. (2012). *Hawaii Papaya Utilization and Price Down*. Retrieved from http://www.nass.usda.gov/Statistics_by_State/Hawaii/Publications/Fruits_and_Nuts/annpapFF.pdf

Navarro, E., Serrano-Heras, G., Castaño, M. J., & Solera, J. (2015). Real-time PCR detection chemistry. *Clinica Chimica Acta, 439*, 231-250. doi:10.1016/j.cca.2014.10.017

Nygren, J., Svanvik, N., & Kubista, M. (1998). The interactions between the fluorescent dye thiazole orange and DNA. *Biopolymers, 46*(1), 39.

Parasnis, A. S., Gupta, V. S., Tamhankar, S. A., & Ranjekar, P. K. (2000). A highly reliable sex diagnostic PCR assay for mass screening of papaya seedlings. *New Strategies in Plant Improvement, 6*(3), 337-344. doi:10.1023/A:1009678807507

Parasnis, A. S., Ramakrishna, W., Chowdari, K. V., Gupta, V. S., & Ranjekar, P. K. (1999). Microsatellite (GATA)n reveals sex-specific differences in Papaya. *International Journal of Plant Breeding Research, 99*(6), 1047-1052. doi:10.1007/s001220051413

Paull, R. E., Nishijima, W., Reyes, M., & Cavaletto, C. (1997). Postharvest handling and losses during marketing of papaya (Carica papaya L.). *Postharvest Biology and Technology, 11*(3), 165-179. doi:10.1016/S0925-5214(97)00028-8

Peterson, D., Boehm, K., & Stack, S. (1997). Isolation of milligram quantities of nuclear DNA from tomato (Lycopersicon esculentum), A plant containing high levels of polyphenolic compounds. *Plant Molecular Biology Reporter, 15*(2), 148-153. doi:10.1007/BF02812265

Primer Design Guide for PCR :: Learn Designing Primers for PCR. Retrieved from http://www.premierbiosoft.com/tech_notes/PCR_Primer_Design.html

Rieger, M. (2006). *Introduction to fruit crops*. New York: New York : Food Products Press.

Sangwan, N. S., Sangwan, R. S., & Kumar, S. (1998). Isolation of Genomic DNA from the Antimalarial Plant Artemisia annua. *Plant Molecular Biology Reporter, 16*(4), 365-365. doi:10.1023/A:1007545102889

Sharma, K., Lavanya, M., & Anjaiah, V. (2000). A method for isolation and purification of peanut genomic DNA suitable for analytical applications. *Plant Molecular Biology Reporter, 18*(4), 393-393. doi:10.1007/BF02825068

Tel-zur, N., Abbo, S., Myslabodski, D., & Mizrahi, Y. (1999). Modified CTAB Procedure for DNA Isolation from Epiphytic Cacti of the Genera Hylocereus and Selenicereus (Cactaceae). *Plant Molecular Biology Reporter, 17*(3), 249-254. doi:10.1023/A:1007656315275

Thornton, B., & Basu, C. (2011). Real-Time PCR (qPCR) Primer Design Using Free Online Software. *Biochemistry and Molecular Biology Education, 39*(2), 145-154. doi:10.1002/bmb.20461

Vanburen, R., Zeng, F., Chen, C., Zhang, J., Wai, C. M., Han, J., . . . Ming, R. (2015). Origin and domestication of papaya Yh chromosome. *Genome research, 25*(4), 524. doi:10.1101/gr.183905.114

Varma, A., Padh, H., & Shrivastava, N. (2007). Plant genomic DNA isolation: An art or a science. *Biotechnology Journal, 2*(3), 386-392. doi:10.1002/biot.200600195

Wang, J., Na, J.-K., Yu, Q., Gschwend, A. R., Han, J., Zeng, F., . . . Ming, R. (2012). Sequencing papaya X and Yh chromosomes reveals molecular basis of incipient sex chromosome evolution. *Proceedings of the National Academy of Sciences of the United States of America, 109*(34), 13710. doi:10.1073/pnas.1207833109

WHO | Micronutrient deficiencies. (n.d., 2013-12-09 15:20:17). *WHO*. Retrieved from http://www.who.int/nutrition/topics/vad/en/

Williams, C. E., & Ronald, P. C. (1994). PCR template-DNA isolated quickly from monocot and dicot leaves without tissue homogenization. *Nucleic acids research, 22*(10), 1917. doi:10.1093/nar/22.10.1917

Xin, Z., Velten, J. P., Oliver, M. J., & Burke, J. J. (2003). High-throughput DNA extraction method suitable for PCR. *BioTechniques, 34*(4), 820.

Xu, W., Bai, W., Guo, F., Luo, Y., Yuan, Y., & Huang, K. (2008). A papaya-specific gene, papain , used as an endogenous reference gene in qualitative and real-time quantitative PCR detection of transgenic papayas. *European Food Research and Technology, 228*(2), 301-309. doi:10.1007/s00217-008-0935-6

Yeh, S.-D. (1988). Control of Papaya Ringspot Virus by Cross Protection. *Plant Dis., 72*(5), 375. doi:10.1094/PD-72-0375

Zhiyong, L., Paul, H. M., Hao, M., Christine, M. A., Makandar, R., Qingyi, Y., . . . Ray, M. (2004). A primitive Y chromosome in papaya marks incipient sex chromosome evolution. *Nature, 427*(6972), 348. doi:10.1038/nature02228

Zhu, X., Li, X., Chen, W., Chen, J., Lu, W., Chen, L., & Fu, D. (2012). Evaluation of New Reference Genes in Papaya for Accurate Transcript Normalization under Different Experimental Conditions (Evaluation of New Reference Genes in Papaya). *PloS one, 7*(8), e44405. doi:10.1371/journal.pone.0044405

Zipper, H., Brunner, H., Bernhagen, J., & Vitzthum, F. (2004). Investigations on DNA intercalation and surface binding by SYBR Green I, its structure determination and methodological implications. *Nucleic acids research, 32*(12), e103.